Solving large scale polynomial convex problems on ℓ_1/nuclear norm balls by randomized first-order algorithms *

Aharon Ben-Tal
Faculty of Industrial Engineering and Management, Technion
Technion city, Haifa 32000, Israel
abental@ie.technion.ac.il

Arkadi Nemirovski
Georgia Institute of Technology, Atlanta, Georgia 30332, USA
nemirovs@isye.gatech.edu

October 26, 2012

Abstract

One of the most attractive recent approaches to processing well-structured large-scale convex optimization problems is based on smooth convex-concave saddle point reformulation of the problem of interest and solving the resulting problem by a fast First Order saddle point method utilizing smoothness of the saddle point cost function. In this paper, we demonstrate that when the saddle point cost function is polynomial, the precise gradients of the cost function required by deterministic First Order saddle point algorithms and becoming prohibitively computationally expensive in the extremely large-scale case, can be replaced with incomparably cheaper computationally unbiased random estimates of the gradients. We show that for large-scale problems with favourable geometry, this randomization accelerates, progressively as the sizes of the problem grow, the solution process. This extends significantly previous results on acceleration by randomization, which, to the best of our knowledge, dealt solely with *bilinear* saddle point problems. We illustrate our theoretical findings by instructive and encouraging numerical experiments.

Key words: convex-concave saddle point problems, large-scale convex programming, first order optimization algorithms, acceleration by randomization.

AMS Subject Classification: 90C06, 90C25, 90C47, 90C52, 68W20.

*Research of both authors was supported by the BSF grant # 2008302. Support of the second author was also supported by NSF grants DMS-0914785 and CMMI-1232623

1 Introduction

The goal of this paper is to develop *randomized* First Order algorithms for solving large-scale "well structured" convex-concave saddle point problems. The background and motivation for our work can be briefly outlined as follows. Theoretically, the entire Convex Programming is within the grasp of polynomial time Interior Point Methods capable to generate high-accuracy solutions at a low iteration count. However, the complexity of an IPM iteration, in general, grows rapidly (as n^3) with the design dimension of the problem, which in numerous applications (like LP's with dense constraint matrices arising in Signal Processing) make IPM's prohibitively time-consuming in the large-scale case. There seemingly is consensus that "beyond the practical grasp of IPM's," one should use the First Order Methods (FOM's) which, under favorable circumstances, allow to get medium-accuracy solutions in (nearly) dimension-independent number of relatively cheap iterations. Over the last decade, there was a significant progress in FOM's; to the best of our understanding, the key to this progress is in discovering a way (Nesterov 2003, see [11]) to utilize problem's structure in order to accelerate FOM algorithms, specifically, to reduce a convex minimization problem $\min_{x\in X} f(x)$ with potentially nonsmooth objective f to a saddle point problem

$$\min_{x\in X}\max_{y\in Y}\phi(x,y), \tag{SP}$$

where ϕ is a $\mathrm{C}^{1,1}$ convex-concave function such that

$$f(x)=\max_{y\in Y}\phi(x,y). \tag{1}$$

The rationale is as follows: when f is nonsmooth (which indeed is the case in typical applications), the (unimprovable in the large-scale case) rate of convergence of FOM's directly applied to the problem of interest $\min_{x\in X} f(x)$ is as low as $O(1/\sqrt{t})$, so that finding a feasible ϵ-optimal solution takes as much as $O(1/\epsilon^2)$ iterations. Utilizing representation (1), this rate can be improved to $O(1/t)$; when X, Y are simple, this dramatic acceleration keeps the iteration's complexity basically intact.

Now, in the original *Nesterov's Smoothing* [11], (1) is used to approximate f by a $\mathrm{C}^{1,1}$ function which is further minimized by Nesterov's optimal algorithm for smooth convex minimization. An alternative is work on (SP) "as it is," by applying to (SP) an $O(1/t)$-converging saddle point FOM, like the Mirror Prox algorithm [8]; in what follows, we further develop this alternative.

When solving (SP) by a FOM, the computational effort per iteration has two components: (a) computing the values of $\nabla\phi$ at $O(1)$ points from $Z=X\times Y$, and (b) "computational overhead," like projecting onto Z. Depending on problem's structure and sizes, any one of these two components can become dominating; the approach we are developing in this paper aimed at the situation where the computational "expenses" related to (a) by far dominate those related to (b), so that the "practical grasp" of the usual – deterministic – saddle point FOMs as applied to (SP) is restricted with the problems where the required number of computations of $\nabla\phi$ (which usually is in the range for hundreds) can be carried out in a reasonable time. An attractive way to lift, to some extent, these restrictions is to pass from the precise values of $\nabla\phi$, which can be prohibitively costly computationally in the large-scale case, to computationally cheap unbiased *random estimates* of these values. This idea (in retrospect, originating from the ad hoc sublinear type matrix game algorithm of Grigoriadis and Khachiyan [3]) has been developed in several papers, see [1, 9, 4, 2, 7], [6, section 6.5.2] and references therein. To the best of our knowledge, for the time being "acceleration via randomization" was developed

solely for the case of saddle point problems with *bilinear* cost function ϕ. The contribution of this paper is in extending the scope of randomization to the case of when ϕ is a *polynomial*.

The main body of this paper is organized as follows. In section 2, we formulate the problem of interest and present the necessary background on our "working horse" — Mirror Prox algorithm. In section 3, we develop a general randomization scheme aimed at producing unbiased random estimates of $\nabla\phi$ for a polynomial ϕ. Theoretical efficiency estimates for the resulting randomized saddle point algorithm are derived in section 4. In section 5, we illustrate our approach by working out in full details two generic examples: optimizing the maximal eigenvalue of a quadratic matrix pencil, and low dimensional approximation of a finite collection of points. We show theoretically (and illustrate by numerical examples) that in both these cases, in a meaningful range of problem's sizes and ϵ, solving problem within accuracy ϵ by randomized algorithm is by far less demanding computationally than achieving the same goal with the best known to us deterministic competitors, and the resulting "acceleration by randomization" goes to ∞ as the problem sizes grow.

2 Situation and Goals

2.1 Problem Statement

Consider the situation as follows: let $X \subset E_x, y \subset E_y$ be convex compact subsets of Euclidean spaces, and let $\phi(x,y) : E := E_x \times E_y \to \mathbf{R}$ be a polynomial of degree d:

$$\phi(z) = \sum_{k=0}^{d} Q_k(\underbrace{z, ..., z}_{k}), \tag{2}$$

where Q_0 is a constant, and for $k > 0$, $Q_k(z^1, ..., z^k)$ is a k-linear symmetric form on E . From now on we assume that $\phi(x,y)$ *is convex-concave on* $X \times Y$, that is, convex in $x \in X$ for fixed $y \in Y$, and concave in $y \in Y$ for fixed $x \in X$.

$$\mathrm{SadVal} = \min_{x\in X} \max_{y\in Y} \phi(x,y). \tag{3}$$

Let

$$\begin{array}{rcll} \mathrm{Opt}(P) & = & \min\limits_{x\in X} \left[\overline{\phi}(x) := \max_{y\in Y} \phi(x,y)\right] & (P) \\ \mathrm{Opt}(D) & = & \max\limits_{y\in Y} \left[\underline{\phi}(y) := \min_{x\in X} \phi(x,y)\right] & (D) \end{array} \tag{4}$$

be the primal-dual pair of convex programs associated with (3), so that $\mathrm{Opt}(P) = \mathrm{Opt}(D)$, let

$$\mathrm{DualityGap}(x,y) = [\overline{\phi}(x) - \mathrm{Opt}(P)] + [\mathrm{Opt}(D) - \underline{\phi}(y)] = \overline{\phi}(x) - \underline{\phi}(y) \tag{5}$$

be the associated duality gap, and, finally, let

$$F(z,y) = \left[F_x(x,y) = \phi'_x(x,y); F_y(x,y) := -\phi'_y(x,y)\right] : Z := X \times Y \to E := E_x \times E_y \tag{6}$$

be the monotone mapping associated with (3). Our ideal goal is, given tolerance $\epsilon > 0$, to find an ϵ-*solution* to (3), i.s., a point $z_\epsilon = (x_\epsilon, y_\epsilon) \in Z$ such that

$$\mathrm{DualityGap}(z_\epsilon) \le \epsilon, \tag{7}$$

whence x_ϵ is a feasible ϵ-optimal solution to (P), while y_ϵ is a feasible ϵ-optimal solution to (D). We intend to achieve this goal by utilizing *randomized* First Order saddle point algorithm, specifically, Stochastic Mirror Prox method (SMP) [4].

2.2 Background on Stochastic Mirror Prox algorithm

The setup for SMP as applied to (3) is given by

- A norm $\|\cdot\|$ on the subspace
$$L[Z] := \mathrm{Lin}(Z - Z)$$
in the embedding space $E := E_x \times E_y$ of the domain $Z := X \times Y$ of the saddle point problem. The (semi)norm on E conjugate to $\|\cdot\|$ is denoted by $\|\cdot\|_*$:
$$\|\zeta\|_* = \max_{z \in L[Z]} \{\langle \zeta, z\rangle : \|z\| \le 1\};$$

- a *distance-generating function* (d.g.-f.) $\omega(z) : Z \to \mathbf{R}$ which should be convex and continuously differentiable on Z, should admit continuous on $Z^o := \{z \in Z : \partial\omega(z) \neq \emptyset\}$ selection $\omega'(z)$ of subgradients and should be *compatible* with $\|\cdot\|$, meaning strong convexity of $\omega(\cdot)$, modulus 1, w.r.t. $\|\cdot\|$:
$$\langle \omega'(z) - \omega'(z'), z - z'\rangle \ge \|z - z'\|^2 \ \forall (z, z' \in Z).$$

A SMP setup induces several important entities, specifically

- ω-center $z_\omega := \mathrm{argmin}_{z \in Z} \omega(z)$ of Z;
- *Bregman distance* $V_z(w) := \omega(w) - \omega(z) - \langle \omega'(z), w - z\rangle$, where $z \in Z^o$ and $w \in Z$. By strong convexity of ω, we have $V_z(w) \ge \frac{1}{2}\|w - z\|^2$;
- *ω-radius* $\Omega := \sqrt{2[\max_Z \omega(\cdot) - \min_Z \omega(\cdot)]}$; noting that $\frac{1}{2}\|w - z_\omega\|^2 \le V_{z_\omega}(w) \le \omega(w) - \omega(z_\omega)$, we conclude that
$$\forall (w \in Z) : \|w - z\| \le \Omega; \tag{8}$$
- *Prox-mapping* $\mathrm{Prox}_z(\xi)$, $z \in Z^o$, $\xi \in E$, defined as
$$\mathrm{Prox}_z(\xi) = \mathrm{argmin}_{w \in Z} [\langle \xi, w\rangle + V_z(w)] = \mathrm{argmin}_{w \in Z} [\langle \xi - \omega'(z), w\rangle + \omega(w)]$$

As applied to (3), SMP operates with *Stochastic Oracle* representation of the vector field F associated with the problem. A *Stochastic Oracle* is a procedure ("black box") which, at t-th call, a point z_t being the input, returns the random vector
$$g(z_t, \xi_t) = F(z_t) + \Delta(z_t, \xi_t) \in E$$
where $\Delta(\cdot, \cdot)$ is a deterministic function, and $\xi_1, \xi_2, ...$ is a sequence of i.i.d. "oracle noises." The SMP algorithm is the recurrence

$$\begin{array}{lrcl}
\text{initialization:} & z_1 & = & z_\omega; \\
\text{search points:} & z_t & \mapsto & w_t = \mathrm{Prox}_{z_t}(\gamma_t g(z_t, \xi_{2t-1})) \mapsto z_{t+1} = \mathrm{Prox}_{z_t}(\gamma_t g(w_t, \xi_{2t})) \mapsto ... \\
\text{approximate solutions:} & z^t & = & (x^t, y^t) = [\sum_{\tau=1}^t \gamma_\tau]^{-1} \sum_{\tau=1}^t \gamma_\tau w_\tau
\end{array} \tag{9}$$

where $\gamma_t > 0$ are deterministic stepsizes.

The main results on SMP we need are as follows (see the case $M = \mu = 0$ of [4, Corollary 1]):

Theorem 2.1 *Assume that $\mathcal{L} < \infty$ and $\sigma < \infty$ are such that*

$$\begin{array}{ll}(a) & \|F(z) - F(z')\|_* \le \mathcal{L}\|z - z'\| \ \forall z, z' \in Z \\ (b) & \mathbf{E}_\xi\{\Delta(z,\xi)\} = 0 \ \forall z \in Z \\ (c) & \mathbf{E}_\xi\{\|\Delta(z,\xi)\|_*^2\} \le \sigma^2 \ \forall z \in Z\end{array} \tag{10}$$

Then for every $t = 1, 2, ...$ the t-step SMP with constant stepsizes

$$\gamma_\tau = \min\left[\frac{1}{\sqrt{3}\mathcal{L}}, \frac{\Omega}{\sqrt{7}\sigma\sqrt{t}}\right], \ 1 \le \tau \le t \tag{11}$$

ensures that

$$\mathbf{E}\{\mathrm{DualityGap}(x^t, y^t)\} \le K_t := \max\left[\frac{2\Omega^2\mathcal{L}}{t}, \frac{6\Omega\sigma}{\sqrt{t}}\right]. \tag{12}$$

In addition, strengthening (10.c) *to*

$$\mathbf{E}_\xi\{\Delta(z,\xi)\} = 0, \ \mathbf{E}\{\exp\{\|\Delta(z,\xi)\|_*^2/\sigma^2\}\} \le \exp\{1\} \tag{13}$$

we have an exponential bound on large deviations: for every $\Lambda > 0$, we have

$$\mathrm{Prob}\left\{\mathrm{DualityGap}(x^t, y^t) > K_t + \Lambda\frac{7\Omega\sigma}{2\sqrt{t}}\right\} \le \exp\{-\Lambda^2/3\} + \exp\{-\Lambda t\}. \tag{14}$$

3 Randomization

Problem (3) by itself is a fully deterministic problem; with "normal" representation of the polynomial $\phi(x, y)$ (e.g., by list of its nonzero coefficients), a precise ($\sigma = 0$) deterministic oracle for F is available; utilizing this oracle, a solution of accuracy ϵ is obtained in $O(1)\Omega^2\mathcal{L}/\epsilon$ iterations, with computational effort per iteration dominated by the necessity to compute the values of F at two points and the values of two prox-mappings. When Z is "simple enough," the complexity of the second of these two tasks – computing prox-mappings – is a tiny fraction of the complexity of precise computation of the values of F. Whenever this is the case, it *might* make sense to replace the precise values F (which can be very costly in the large-scale case) with computationally cheap unbiased random estimates of these values. This is the option we intend to investigate in this paper. We start with a general description of the randomization we intend to use.

Observe, first, that

$$F(z) = D\nabla\phi(z)$$

where $D = \mathrm{Diag}\{\mathrm{Id}_x, -\mathrm{Id}_y\}$, Id_x and Id_y being the identity mappings on E_x and E_y, respectively. Now, representing the polynomial $\phi(z)$ as

$$\phi(z) = \sum_{k=0}^{d} Q_k(\underbrace{z, ..., z}_{k}), \tag{15}$$

where $Q_k(z^1, ..., z^k)$ is a symmetric k-linear form on E, we have

$$\langle F(z), h\rangle = \langle D\nabla\phi(z), h\rangle = \langle\nabla\phi(z), Dh\rangle = \sum_{k=1}^{d} kQ_k(Dh, \underbrace{z, ..., z}_{k-1}) \tag{16}$$

Now assume that we can associate with every $z \in Z$ a probability distribution P_z on E such that

$$\int \xi dP_z(\xi) = z \ \forall z \in E. \tag{17}$$

In order to get an unbiased estimate of $F(z)$, one can act as follows:

- given z, draw $d-1$ independent samples $z^i \sim P_z$, $i = 1, ..., d-1$
- compute the linear form $G = G[z^1, ..., z^{d-1}]$ on E given by

$$\forall h \in E : \langle G, h \rangle = \sum_{k=1}^{d} kQ_k(Dh, z^1, z^2, ..., z^{k-1}). \tag{18}$$

  thus ensuring that

$$\mathbf{E}_{(z^1,...,z^{d-1})\sim P_z \times ... \times P_z}\{G[z^1, ..., z^{d-1}]\} = F(z) \ \forall z \in Z. \tag{19}$$

Note that we can represent a random variable distributed according to P_z as a deterministic function of z as a standard random variable ξ uniformly distributed on $[0, 1]$, which makes G a deterministic function of z and $\xi \sim \text{Uniform}[0, 1]$, as required by our model of a Stochastic Oracle.

Observe that for a general-type convex-concave polynomial $\phi(x, y)$ of degree d, precise deterministic computation of $F(z)$ is as suggested by (18) *with P_z being the unit mass sitting at the singleton z*, that is, with $z^1 = ... = z^{d-1} = z$. It follows that *if the distributions P_z, for every $z \in Z$ are such that computing the vectors g_k of coefficients of the linear forms $Q_k(h, z^1, ..., z^{k-1})$ of $h \in E$ is much cheaper than the similar task for the linear forms $Q_k(h, z, ..., z)$ for a "general position" $z \in Z$, then computing the unbiased estimate $G = G[z^1, ..., z^{d-1}]$ of $F(z)$ is much cheaper computationally than the precise computation of $F(z)$*, so that there are chances for the outlined randomization to reduce the overall complexity of computing ϵ-solution to (3). Let us look at two simple preliminary examples:

Example 1 ["Scalar case"]: E is just the space $\mathbf{R}^n$ of n-dimensional vectors, and the k-linear forms $Q_k(\cdot)$ are given by lists of their nonzero coefficients. In this case, we can specify P_z as follows:

Given $z \in E = \mathbf{R}^n \backslash \{0\}$, let P_z be the discrete probability distribution supported on the set $\{f^i = \text{sign}(z_i)\|z\|_1 e_i\}_{i=1}^n$, where e^i are the standard basic orths in E, with the probability mass of f^i equal to $|z_i|/\|z\|_1$; when $z = 0$, let P_z be the unit mass sitting at the origin. We clearly have $\mathbf{E}_{f\sim P_z}\{f\} = z$, and all realizations of $f \sim P_z$ are extremely sparse — with at most one nonzero entry. Now, in order to generate $f \sim P_z$, we need preprocessing of $O(1)n$ a.o. aimed to compute $\|z\|_1$ and the "cumulative distribution" $s_0 = 0$, $s_i = \|z\|_1^{-1} \sum_{j=1}^{i} |z_j|$, $i = 1, ..., n$. With this cumulative distribution at hand, to draw a sample $f \sim P_z$ takes just $O(1) \ln(n)$ a.o.: we draw at random a real α uniformly distributed in $[-1, 1]$ (which for all practical purposed is jus $O(1)$ a.o.), find by bisection the smallest $i \in \{1, ..., n\}$ such that $\alpha \leq s_i$ ($O(1) \ln(n)$ a.o.) and return the vector $f = \text{sign}(z_i)\|z\|_1 e_i$ ($O(1)$ a.o.). Thus, generating $z^1, ..., z^{d-1}$ costs $O(1)[n + d \ln(n)]$ a.o. Now, with our "ultimately sparse" $z^1, ..., z^{d-1}$, computing the coefficients of the linear form $Q_k(h, z^1, ..., z^{k-1})$ of h takes at most $n(d+\mathcal{C})$ a.o., where $\mathcal{C}$ is the (upper bound

on) the cost of extracting the coefficient of the k-linear symmetric form, given its "address." The bottom line is that the complexity of computing $G[z^1, .., z^{d-1}]$ is

$$\mathcal{C}_r[P] = O(1)d\left[n + d\ln(n) + n(d + \mathcal{C})\right] = O(1)[d^2 n + dn\mathcal{C}] \text{ a.o.}$$

On the other hand, computing $F(z)$ exactly costs something like

$$\mathcal{C}_d[P] = O(1)[n + \sum_{k=1}^{d} kN_k\mathcal{C}] \text{ a.o.}$$

where N_k is the total number of nonzero coefficients in $Q_k(\cdot, ..., \cdot)$. Assuming that $d = O(1)$, we see that *unless all Q_k are pretty sparse – just with $N_k = O(n)$ nonzero coefficients, mimicking unbiased Stochastic Oracle takes by orders of magnitude less computations than precise deterministic computation of $F(z)$.*

4 Complexity Analysis

The discussion in the previous section demonstrates that in some interesting cases unbiased random estimates of the vector field F associated with (3) are significantly cheaper computationally than the precise values of F. This does not mean, however, that in all these cases randomization is profitable — it well may happen that as far as the overall complexity of ϵ-solution is concerned, expensive high-quality local information is better than cheap low quality one. We intend to analyze the situation in the regime when the degree d of the polynomial ϕ is a small integer formally treated as $O(1)$, so that we ignore the details of dependence of hidden factors in the estimates to follow on d.

4.1 Preliminaries

Standing Assumptions. Observe that

$$L[Z] := \mathrm{Lin}(Z - Z) = \mathrm{Lin}(X - X) \times \mathrm{Lin}(Y - Y) = L[X] \times L[Y]. \tag{20}$$

Now, the sets

$$X^s = \frac{1}{2}[X - X],\ Y^s = \frac{1}{2}[Y - Y],\ Z^s = \frac{1}{2}[Z - Z] = X^s \times Y^s$$

are unit balls of certain norms $\|\cdot\|_X$ on $L[X]$, $\|\cdot\|_Y$ on $L[Y]$ and $\|\cdot\|$ on $L[Z]$, with

$$\|(x, y)\| = \max[\|x\|_X, \|y\|_Y],\ x \in L[X], y \in L[Y]. \tag{21}$$

From now on, we make the following

> **Assumption A.** *The just defined norm $\|\cdot\|$ with the unit ball $\frac{1}{2}[Z - Z]$ is the norm used in the SMP setup, while the d.-g.f $\omega(x, y)$ is of the form $\omega_X(x) + \omega_Y(y)$, where $(\|\cdot\|_X, \omega_X(\cdot))$ and $(\|\cdot\|_Y, \omega_Y(\cdot))$ form SMP setups for (X, E_x) and (Y, E_y) respectively*[1].

[1]Note that such a sum indeed is a d.-g.f. fitting the norm $\|\cdot\|$.

Note that

- We have
$$\|[\xi;\eta]\|_* = \|\xi\|_{X,*} + \|\eta\|_{Y,*}, \tag{22}$$
where $\|\cdot\|_{X,*}$ and $\|\cdot\|_{Y,*}$ are the (semi)norms conjugate to $\|\cdot\|_X$, $\|\cdot\|_Y$, respectively. In particular, we have
$$\|F(z)\|_* = \|\nabla\phi(z)\|_*,\ \|F(z) - F(z')\|_* = \|\nabla\phi(z) - \nabla\phi(z')\|_*\ \forall z, z' \in E. \tag{23}$$

- The ω-radius Ω of Z is
$$\Omega = \sqrt{\Omega_X^2 + \Omega_Y^2}, \Omega_X = \sqrt{2[\max_{x\in X}\omega_X(x) - \min_{x\in X}\omega_X(x)]}, \Omega_Y = \sqrt{2[\max_{y\in Y}\omega_Y(y) - \min_{y\in Y}\omega_Y(y)]} \tag{24}$$

Scale factor. When speaking about complexity of finding ϵ-solution, we will express this complexity in terms of the *relative accuracy* $\nu = \epsilon/\mathbf{V}$, where the *scale factor* $\mathbf{V}$ is defined as follows. Let $\widehat{Z}$ be the convex hull of $\{0\} \cup Z$, and let

$$\widehat{\phi}(z) = \phi(z) - \phi(0) - \langle\phi'(0), z\rangle = \sum_{k=2}^{d} Q_k(z, ..., z).$$

We set
$$\mathbf{V} = \mathbf{V}_Z[\phi] := \max_{z\in\widehat{Z}}\widehat{\phi}(z) - \min_{z\in\widehat{Z}}\widehat{\phi}(z). \tag{25}$$
The importance of this scale factor in our contents stems from the following simple observation (see also Lemma 4.2 below):

Lemma 4.1 *For properly chosen positive real $C^{(1)}$ depending solely on d, for all k, $2 \le k \le d$ and all collections $z^1, ..., z^k$ of vectors from $L[\widehat{Z}]$ one has*

$$|Q_k(z^1, ..., z^k)| \le C^{(1)}\mathbf{V}\prod_{i=1}^{k}\|z^i\|_{\widehat{Z}} \tag{26}$$

In particular, the vector field $F(z)$ associated with (3) *satisfies* (10.a) *with*

$$\mathcal{L} = C^{(1)}\mathbf{V}\sum_{k=2}^{d} k(k-1)2^{k-2} := C^{(2)}\mathbf{V}, \tag{27}$$

where $C^{(2)}$ depends solely on d.

For proof, see Appendix.

An immediate question related to the definition of the scaling factor is: as defined, a "shift of the problem by $a \in E$" – a simple substitution of variables $z = w - a$ – changes the factor and thus the complexity estimates, although such a substitution leaves the problem "the same." The answer is as follows: while the "shift option" should be kept in mind, such a shift changes the Stochastic Oracle as given by (18). Indeed, this oracle is defined in terms of the *homogeneous components in the Taylor decomposition of $\phi(\cdot)$ taken at the origin*, and this is why the origin is participating in the description of $\widehat{Z}$ and thus of $\mathbf{V}$. Shifting the origin, we, in general, change

the $\mathcal{SO}$ [2], and thus there is nothing strange that our scaling of the accuracy (and thus – the efficiency estimates) corresponding to a given Z and a given (implicitly participating in (18)) $\mathcal{SO}$ are not translation-invariant.

4.2 Complexity Analysis

Preliminaries. From now on we assume that as applied to (3), SMP utilizes Stochastic Oracle $\mathcal{SO}$ given according to (18) by a family of probability distributions $\mathcal{P} = \{P_z : z \in Z\}$ on E satisfying (17). From now on, we make the following

> **Assumption B.** *For some $\rho \geq 0$, all distributions P_z, $z \in Z$, are supported on the set $Z + 2\rho Z^s \subset \mathrm{Aff}(Z)$, where $Z^s = \frac{1}{2}[Z - Z]$ and $\mathrm{Aff}(Z)$ is the affine hull of Z.*

In particular, when P_z is supported on Z for all $z \in Z$ ("proper case"), Assumption B is satisfied with $\rho = 0$.

It is time now to note that the $\mathcal{SO}$ we have developed so far gives rise to a *parametric family* of Stochastic Oracles, specifically, as follows. First of all, our basic $\mathcal{SO}$ in fact can be "split" into two Stochastic Oracles, $\mathcal{SO}^x$ and $\mathcal{SO}^y$, providing estimates of the x- and the y-components F_x, F_y of $F(z) = [F_x(z); F_y(z)]$: the estimates

$$\begin{array}{rcl} E_x \ni G_x & = & G_x[z^1, ..., z^{d-1}] : \forall \xi \in E_x : \langle G_x, \xi \rangle = \sum_{k=1}^d k Q_k([\xi; 0], z^1, ..., z^{k-1}), \\ E_y \ni G_y & = & G_y[z^1, ..., z^{d-1}] : \forall \eta \in E_y : \langle G_y, \eta \rangle = -\sum_{k=1}^d k Q_k([0; \eta], z^1, ..., z^{k-1}). \end{array}$$

Here, as above, $z^1, ..., z^{d-1}$ are, independently of each other, sampled from P_z. Now, given two positive integers k_x, k_y, we can "recombine" our "partial stochastic oracles" $\mathcal{SO}^x$, $\mathcal{SO}^y$ into a new Stochastic Oracle $\mathcal{SO}_{k_x,k_y}$ as follows: in order to generate a random estimate of $F(z)$ given $z \in Z$, we generate $(d-1)\max[k_x, k_y]$ independent samples $z_\tau^k \sim P_z$, $1 \leq k \leq d-1$, $1 \leq \tau \leq k_{xy} := \max[k_x, k_y]$ and then set

$$g = G_z^{k_x,k_y}\left[\{z_\tau^k\}_{\substack{1 \leq k \leq d-1, \\ 1 \leq \tau \leq k_{xy}}}\right] = \left[\frac{1}{k_x}\sum_{\tau=1}^{k_x} G_x[z_\tau^1, ..., z_\tau^{d-1}]; \frac{1}{k_y}\sum_{\tau=1}^{k_y} G_y[z_\tau^1, ..., z_\tau^{d-1}]\right]. \qquad (28)$$

In the sequel, we refer to k_x and k_y as the *x- and y- multiplicities* of the Stochastic Oracle $\mathcal{SO}_{k_x,k_y}$.

We will make use of the following

Lemma 4.2 *Under Assumptions A, B, for all positive integer multiplicities k_x, k_y, $\mathcal{SO}_{k_x,k_y}$ ensures validity of (10.b), same as the validity of (13) with*

$$\sigma = C^{(3)}\mathbf{V}(1+\rho)^{d-1}\left[\min[1, \Omega_X/\sqrt{k_x}] + \min[1, \Omega_Y/\sqrt{k_y}]\right], \qquad (29)$$

where $C^{(3)}$ depends solely on d.

For proof, see Appendix.

We have arrived at the following

[2] For example, with $\phi(x, y) \equiv x^3$, the oracle (18) is $G = [3x^1x^2; 0]$, $z^i = [x^i; 0] \sim P_z$. Substituting $x = 1 + h$, carrying out the construction of the $\mathcal{SO}$ "in h-variable" and translating the result back to x-variable, the resulting $\mathcal{SO}$ turns out to be $G = [3x^1x^2 + 3x^1 - 3x^2; 0]$, which is not the oracle we started with.

Theorem 4.1 *Let $t \geq 1$ be given, let Assumptions A, B be satisfied, and let problem* (3) *be solved by t-step SMP utilizing $\mathcal{SO}_{k_x,k_y}$, with the parameters $\mathcal{L}$, σ underlying the stepsize policy* (11)*'s given by* (27), (29)*. Then, for some C depending solely on d,*

$$(a) \quad \mathbf{E}\{\mathrm{DualityGap}(x^t, y^t)\} \leq K(t) := C\left[\frac{\Omega_X^2+\Omega_Y^2}{t} + \frac{\sqrt{\Omega_X^2+\Omega_Y^2}(1+\rho)^{d-1}\vartheta}{\sqrt{t}}\right]\mathbf{V}$$
$$\vartheta = \min[1, \Omega_X/\sqrt{k_x}] + \min[1, \Omega_Y/\sqrt{k_y}];$$
$$(b) \quad \mathrm{Prob}\left\{\mathrm{DualityGap}(x^t, y^t) > K(t) + C\Lambda\frac{\sqrt{\Omega_X^2+\Omega_Y^2}(1+\rho)^{d-1}\vartheta\mathbf{V}}{\sqrt{t}}\right\} \leq \exp\{-\Lambda^2/3\} + \exp\{-\Lambda t\}.$$
$$\forall \Lambda > 0. \tag{30}$$

5 Illustrations

We illustrate the proposed approach by two examples. The first of them is of a purely academic nature, the second can pretend to be of some practical interest. when selecting the examples, our major goal was to illustrate randomization schemes different from the one in Example 1.

5.1 Illustration I: minimizing the maximal eigenvalue of a quadratic matrix pencil

The problem we are interested in is as follows: We are given a symmetric matrix quadratically depending on the "design variables" $x_1, ..., x_J$ which themselves are matrices:

$$\mathcal{A}(x) = \sum_{i=1}^{I}\left[a_i^T x_{j(i)}^T q_i x_{j(i)} a_i + b_i^T x_{j(i)} c_i + c_i^T x_{j(i)}^T b_i\right] + d \in E_y := \mathbf{S}^m, \tag{31}$$

where

- $\mathbf{S}^m$ is the space of $m \times m$ symmetric matrices equipped with the Frobenius inner product,
- $x = \{x_j \in \mathbf{R}^{m_j \times n_j}\}_{j=1}^J$ is a collection of variable matrices which we treat as a block-diagonal rectangular matrix with diagonal blocks x_j, $1 \leq j \leq J$. We denote the linear space of all these matrices by E_x and equip it with the Frobenius inner product;
- $j(i) \in \{1, ..., J\}$, $1 \leq i \leq I$, are given integers,
- $\{a_i, b_i, c_i, q_i\}_{i=1}^I$, d are data matrices of appropriate sizes and structures:

  $$a_i, c_i \in \mathbf{R}^{n_{j(i)} \times m},\ b_i \in \mathbf{R}^{m_{j(i)} \times m},\ q_i \in \mathbf{S}^{m_{j(i)}},\ d \in \mathbf{S}^m;$$

  in addition, we assume that *all q_i are positive semidefinite*, and that the values $j(i)$, $1 \leq i \leq I$, cover the entire range $1 \leq j \leq J$, meaning that every one of the blocks x_j indeed participates in $\mathcal{A}(\cdot)$.

For a matrix $a \in \mathbf{R}^{p \times q}$, let $\sigma(a) = [\sigma_1(a); ...; \sigma_{\min[p,q]}(a)]$ be the vector of singular values of a arranged in the non-ascending order, and let $\|a\|_{\mathrm{nuc}} = \|\sigma(a)\|_1$ be the nuclear norm of a. for a symmetric matrix a, let $\lambda_{\max}(a)$ be the maximal eigenvalue of a. Finally, let

$$X = \{x \in E_x : \|x\|_{\mathrm{nuc}} \leq 1\}.$$

Our goal is to solve the optimization problem

$$\mathrm{Opt} = \min_{x \in X} \{\lambda_{\max}(\mathcal{A}(x))\}, \tag{32}$$

Denoting by Y the standard *spectahedron* in $\mathbf{S}^m$:

$$Y = \{y \in \mathbf{S}^m : y \succeq 0, \mathrm{Tr}(y) = 1\}$$

and observing that $\lambda_{\max}(a) = \max_y\{\mathrm{Tr}(ay) : y \in Y\}$, we can convert the problem of interest into the saddle point problem as follows:

$$\mathrm{Opt} = \min_{x \in X} \max_{y \in Y} [\phi(x, y) := \mathrm{Tr}(y\mathcal{A}(x))]. \tag{33}$$

From $q_i \succeq 0$, $i \leq I$, and the fact that $g + y \succeq 0$ for all $y \in Y$ it immediately follows that the restriction of ϕ on $y \in Y$ is convex in $x \in E_x$; as a function of y, ϕ is a convex-concave on $X \times Y$ polynomial of degree $d = 3$. The monotone mapping (6) associated with (33) is

$$\begin{array}{rcl} F_x(x, y) & = & 2\mathrm{Diag}\{\sum_{i:j(i)=j}[q_i x_j a_i y a_i^T + b_i y c_i^T], 1 \leq j \leq J\} \in E_x, \\ F_y(x, y) & = & -\mathcal{A}(x), \end{array} \tag{34}$$

Now let us apply to (33) the approach we have developed so far.

A. First, let us fix the setup for SMP. We are in the situation when $X^s := \frac{1}{2}[X - X]$ is X – the unit ball of the nuclear norm on E_x; thus, $\|\cdot\|_X$ is the nuclear norm on E_x. The set $Y^s = \frac{1}{2}[Y - Y]$ clearly is contained in the unit nuclear norm ball of $\mathbf{S}^m$ and contains the concentric nuclear norm ball of radius 1/2, meaning that $\|\cdot\|_Y$ is within factor 2 of the nuclear norm:

$$2\|y\|_{\mathrm{nuc}} \geq \|y\|_Y \geq \|y\|_{\mathrm{nuc}} \ \forall y \in \mathbf{S}^m = E_y.$$

The best, within $O(1)$ factors, known so far under circumstances choice of the d.-g.f.'s is (see [6, section 5.7.1] or Propositions A.3, A.2 in Appendix)

$$\begin{array}{rcl} \omega_X(x) & = & O(1)\ln(n)\sum_{\ell=1}^n \sigma_\ell^{q(n)}(x),\ n = \sum_{j=1}^J \min[m_j, n_j],\ q(n) = \frac{1}{2\ln(n)}, \quad {}^{3)} \\ \omega_Y(y) & = & O(1)\ln(m)\sum_{\ell=1}^m \sigma_\ell^{p(m)}(y),\ p(m) = \frac{1}{2\ln(m)}, \end{array} \tag{35}$$

with explicitly given absolute constants $O(1)$. This choice is reasonably good in terms of the values of the corresponding radii of X, Y which turn to be "quite moderate:"

$$\Omega_X \leq O(1)\sqrt{\ln(n)},\ \Omega_Y \leq O(1)\sqrt{\ln(m)}. \tag{36}$$

Note that the efficiency estimate (30) says that we are interested in as small values of Ω_X, Ω_Y as possible. At the same time, it is immediately seen that if $\omega(\cdot)$ is a d.-g-.f. for Z compatible with the norm generated by Z (i.e., with the unit ball $Z^s = \frac{1}{2}[Z - Z]$, then the ω-radius of Z is *at least* $O(1)$, so that Ω_X, Ω_Y are "nearly as good" as the could be.

The outlined d.-g.f.'s are also the best known under circumstances in terms of the computational complexity of the associated prox-mapping; it is easily seen that this complexity is dominated by the necessity to carry out singular value decomposition of a matrix from E_x (which takes $O(\sum_j m_j n_j \min[m_j, n_j])$ a.o.) and eigenvalue decomposition of a matrix from $\mathbf{S}^m$ ($O(m^3)$ a.o.), see below.

[3] To avoid trivial situations, we assume from now on that $m > 1$, $n > 1$.

B. With our approach, the "basic" option when solving (33) is to use the deterministic version of SMP, i.e., to use as P_z the unit mass sitting at z. The corresponding efficiency estimate can be obtained from (30) by setting $k_x = k_y = \infty$; taking into account (36), the resulting estimate says that a solution to (33) of a given accuracy $\epsilon \leq \mathbf{V}$ will be found in course of

$$N_{\rm d}(\epsilon/\mathbf{V}) = O(1)\ln(mn)\mathbf{V}/\epsilon \tag{37}$$

iterations. Now let us evaluate the arithmetic complexity of an iteration. From the description of the algorithm it is clear than the computational effort at an iteration is dominated by the necessity to compute exactly $O(1)$ values of the monotone mapping (34) and of $O(1)$ prox mappings. To simplify evaluating the computational cost of an iteration, assume from now on that we are in the *simple case*:

$$m_j = n_j = \nu,\ 1 \leq j \leq J.$$

In this case, computing $O(1)$ values of the prox mapping costs

$$\mathcal{C}_{\rm prox} = O(1)[m^3 + J\nu^3] \text{ a.o.}$$

> Indeed, with our $\omega_X(\cdot)$, computing the x-component of prox mapping reduces to solving the optimization problem $\min_{v\in E_x, \|v\|_{\rm nuc}\leq 1}[\sum_{\ell=1}^n \sigma_\ell^p(v) - {\rm Tr}(g^T v)]$ with a given $p \in (1,2]$ and a given $g \in E_x$. To solve the problem, we compute the singular value decompositions of all diagonal blocks g_j in g, this getting a representation $g = U{\rm Diag}\{\gamma\}V^T$ with block-diagonal orthogonal matrices U, V, which takes $O(1)J\nu^3$ a.o. It is immediately seen that the problem admits an optimal solution v of the same structure as g: $v = U{\rm Diag}\{\upsilon\}V^T$. Specifying υ reduces to solving the convex optimization problem
>
> $$\min_{\upsilon\in\mathbf{R}^n:\|\upsilon\|_1\leq 1}\left[\sum_j [|\upsilon_j|^p + \gamma_j\upsilon_j\right];$$
>
> this convex problem with separable objective and a single separable constraint clearly can be solved within machine precision in $O(n)$ a.o. Finally, given υ, it takes $O(1)J\nu^3$ operations to compute the x-component $U{\rm Diag}\{\upsilon\}V^T$ of the prox mapping. Thus, the total cost of the x-component of the prox mapping is $O(J\nu^3$ a.o. The situation with computing the y-component of the mapping is completely similar, and the cost of this component is $O(1)m^3$ a.o.

Looking at (34), we see that computing $O(1)$ values of F at a "general position" points x, assuming all the data matrices dense, is

$$\mathcal{C}_F = O(1)\nu m(\nu + m)I \text{ a.o.}$$

As a result, the arithmetic cost of finding ϵ-solution to (33) (and thus – to (32)) by the deterministic version of SMP is

$$\mathcal{C}_{\rm d}(\epsilon) = O(1)\ln(mn)\underbrace{\left[\overbrace{m^3 + J\nu^3}^{\Theta_{\rm prox}} + \overbrace{m\nu(m+\nu)I}^{\Theta_F}\right]}_{\Theta}\frac{\mathbf{V}}{\epsilon} \text{ a.o.} \tag{38}$$

Note that we are not aware of better complexity bounds for large-scale problems (32), at least in the case when in the expression for Θ, the term m^3 is dominated by the sum of other terms.

C. Now let us look whether we can reduce the overall arithmetic cost of ϵ-solution to (32) by randomization. An immediate observation is that the only case when it can happen is the one of $\Theta_F \gg \Theta_{\rm prox}$. Indeed, comparing the efficiency estimates (30) and (37), we conclude that randomization can only increase the iteration cost of ϵ-solution; in order to overweigh the growth in the number of iterations, we need to reduce significantly the arithmetic cost of an iteration, and to this end, this cost, in the deterministic case, should be by far dominated by the cost of computing the values of F (the only component of our computational effort which can be reduced by randomization). Assuming $\Theta_f \gg \Theta_{\rm prox}$, let us look which kind of randomization could be useful in our context. Note that in order for randomization to be useful, the underlying distributions P_z should be supported on the set of those pairs (x, y) for which computing an estimate g of $F(z)$ according to (28) is much cheaper than computing F at a general-type point $(x, y) \in E_x \times E_y$. A natural way to meet this requirement us to use the "matrix analogy" of Example 1, where P_z are supported on the set of low rank matrices. Specifically, in order to get an unbiased estimate of $F(z)$, $z = (x, y) \in X \times Y$, let us act as follows:

1. We compute singular value decomposition $x = U{\rm Diag}\{\sigma(x)\}V^T$ of x and eigenvalue decomposition $y = W{\rm Diag}\{\sigma(y)\}W^T$ of y [4], where U, W are block-diagonal $n \times n$ orthogonal with $\nu \times \nu$ diagonal blocks, and W is an orthogonal $m \times m$ matrix.

2. We specify P_x as the distribution of a random matrix $\xi \in E_x$ with takes the values

$$\|\sigma(x)\|_1 {\rm Col}_j[U]{\rm Col}_j^T[V],\ 1 \le j \le n$$

   with the probabilities $\sigma_j(x)/\|\sigma(x)\|_1$ (when $\sigma(x) = 0$, ξ takes value 0 with probability 1); here ${\rm Col}_j(A)$ denotes j-th column of a matrix A.

3. We specify P_y as the distribution of the random symmetric matrix η which takes values ${\rm Col}_i[W]$, $1 \le i \le m$, with probabilities $\sigma_i(y)$, and specify P_z as the direct product of P_x and P_y.

Observe that the expectation of $\zeta \sim P_z$ is exactly z, and that P_z, $z \in Z = X \times Y$, is supported on Z due to $\|\sigma(x)\|_1 \le 1$, $x \in X$, $\|\sigma(y)\|_1 = 1$, $y \in Y$. In other words, *assumption B is satisfied with* $\rho = 0$.

Note that with the just defined P_z, a realization $\zeta = (\xi, \eta) \sim P_z$ is of very special structure:

$$\xi = u \times v^T,\ u, v \in \mathbf{R}^n,\ \eta = ww^T,\ w \in \mathbf{R}^m; \tag{39}$$

moreover, among the J consecutive ν-dimensional blocks u_j, v_j, $j = 1, ..., J$, of every one of the vectors $u, v \in \mathbf{R}^{n=J\nu}$, all but one blocks are zero, and the nonzero blocks u^j, v^j share a common index j.

[4] Note that the singular values of y are the same as eigenvalues, since $y \succeq 0$ due to $y \in Y$.

It is immediately seen that with the just defined distributions P_z, the unbiased estimate (28) of $F(z)$ is as follows:

$$\begin{array}{rcl} G_x^{k_x} & = & \frac{1}{k_x}\sum_{\ell=1}^{k_x}\mathrm{Diag}\Big\{\sum_{i:j(i)=j}\big[q_i u_j^{2\ell-1}[v_j^{2\ell-1}]^T a_i w^{2\ell}[w^{2\ell}]^T a_i^T + q_i u_j^{2\ell}[v_j^{2\ell}]^T a_i w^{2\ell-1}[w^{2\ell-1}]^T a_i^T \\ & & \qquad + b_i w^{2\ell-1}[w^{2\ell-1}]^T c_i^T\big], j=1,...,J\Big\} \\ G_y^{k_y} & = & -d - \frac{1}{k_y}\sum_{\ell=1}^{k_y}\sum_{i=1}^{I}\Big[\frac{1}{2}a_i^T v_{j(i)}^{2\ell-1}[u_{j(i)}^{2\ell-1}]^T q_i u_{j(i)}^{2\ell}[v_{j(i)}^{2\ell}]^T a_i + \frac{1}{2}a_i^T v_{j(i)}^{2\ell}[u_{j(i)}^{2\ell}]^T q_i u_{j(i)}^{2\ell-1}[v_{j(i)}^{2\ell-1}]^T \\ & & \qquad + b_i^T u_{j(i)}^{2\ell-1}[v_{j(i)}^{2\ell-1}]^T c_i + c_i^T v_{j(i)}^{2\ell-1}[u_{j(i)}^{2\ell-1}]^T b_i\Big] \end{array} \tag{40}$$

where the collections

$$\zeta^\ell = \left([u_1^\ell;...;u_J^\ell][v_1^\ell;...;v_J^\ell]^T, w^\ell[w^\ell]^T\right),\ \ell = 1,...,2\max[k_x,k_y]$$

are independently of each other drawn from P_z.

It is immediately seen that the arithmetic cost of computing (G_x, G_y) given $z=(x,y)$ is comprised of the components as follows:

1. "setup cost" – one of computing singular value decomposition of x and eigenvalue decomposition of y [5] ($O(1)(m^3+J\nu^3)$ a.o.) plus the cost of computing the "cumulative distributions" $S_j(x) = \sum\tau = 1^j \sigma_\tau(x)/\|\sigma(x)\|_1$, $1\le j\le J\nu$, $S_i(y) = \sum_{\tau=1}^i \sigma_i(y)$ ($O(1)(m+J\nu)$ a.o.).

2. After the setup cost is paid, for every ℓ
   — generating ζ^ℓ costs $O(1)(\ln(m)+\ln(J\nu)+m+\nu)$ a.o.,
   — computing the contribution of ζ^ℓ to G_x costs no more than $O(1)I\nu(m+\nu)$ a.o. (look at (40) and take into account that the vectors $u^{2\ell-1}$, $v^{2\ell-1}$ have a single nonzero ν-dimensional block each), and this cost should be paid k_x times;
   — computing the contribution of ζ^ℓ to G_y costs at most $O(1)(m+\nu)^2K$ a.o., where $K = \max_{1\le j\le J}\mathrm{Card}\{i: j(i)=j\}$ (the same argument as above), and this cost can be paid at most $2k_y$ times.

Thus, the cost of computing $(G_x^{k_x}, G_y^{k_y})$ is

$$O(1)\left(m^3+J\nu^3+k_x\nu(m+\nu)I+k_y(m+\nu)^2K\right)\ \text{a.o.},\ K=\max_{1\le j\le J}\mathrm{Card}\{i:j(i)=j\}. \tag{41}$$

To simplify the analysis to follow, assume from now on that $I=J$ and $j(\cdot)$ is one-to-one. In this case $K=1$ and the cost of an iteration is

$$O(1)\left(m^3+J\nu^3+k_x\nu(m+\nu)J+k_y(m+\nu)^2\right)\ \text{a.o.} \tag{42}$$

Now let us evaluate the overall complexity of finding, with confidence $1-\delta$, $\delta\ll 1$, an ϵ-solution by the randomized SMP. We assume from now on that $\epsilon\le\mathbf{V}$ (otherwise the problem is trivial, since $\mathrm{DualityGap}(z)\le\mathbf{V}$ for every $z\in X\times Y$). For the sake of simplicity, we restrict ourselves

[5] In fact, this cost is nonexisting: by construction of the method, the points z where one needs to evaluate F are the values of already computed prox-mappings; according to how we compute these values (see above), they go together with their singular value/eigenvalue decompositions.

with the case of $k_x = k_y = 1$. Invoking the efficiency estimate (30) and taking into account (36) and the fact that we are in the situation of $\rho = 0$, the number t of iterations which results in $\mathrm{DualityGap}(x^t, y^t) \le \epsilon$ is bounded from above by

$$N_{\mathrm{r},\delta}(\epsilon) = O(1)\ln(mn)\ln(1/\delta)(\mathbf{V}/\epsilon)^2,$$

meaning that the iteration count now is nearly square of the one for the deterministic algorithm, see (37). Taking into account (42), the overall complexity of achieving our goal with the randomized algorithm does not exceed

$$\mathcal{C}_{\mathrm{r},\delta}(\epsilon) = O(1)\ln(mn)\ln(1/\delta)\left[m^3 + J\nu^3 + (m+\nu)(m+\nu J)\right](\mathbf{V}/\epsilon)^2 \text{ a.o.}$$

The ratio of this quantity and the "deterministic complexity" (see (38) and take into account that we are in the case of $I = J$) is

$$\mathcal{R} = \frac{\mathcal{C}_{\mathrm{r},\delta}(\epsilon)}{\mathcal{C}_{\mathrm{d}}(\epsilon)} = O(1)\ln(1/\delta)\underbrace{\left[\frac{m^3 + \nu^3 J + (m+\nu)(m+\nu J)}{m^3 + \nu^3 J + m\nu(m+\nu)J}\right]}_{r}\cdot\frac{\mathbf{V}}{\epsilon}.$$

It is immediately seen that *when $\mathbf{V}/\epsilon$ and δ are fixed, and m, ν, J vary in such a way that $m, n = \nu J$ go to ∞ and ν/m, m/n go to 0*, r goes to 0, meaning that eventually the randomized algorithm outperforms its deterministic competitor, and the "performance ratio" goes to ∞ as the sizes m, n of the problem grow.

Numerical illustration. In the experiment we are about to describe, the sizes of problem (32) were selected as

$$m = 300,\ \mu = \nu = 2,\ I = J = 5000,\ j(i) \equiv i$$

which results in $\dim x = 20000$, $\dim y = 45150$. The data matrices $q_i \succeq 0, a_i, b_i, c_i$ were generated at random and normalized to have spectral norms 1, which ensures $\mathbf{V} \le 1$. A generated instance was processed as follows:

- first, it was solved by the deterministic Mirror Prox algorithm (DMP) with on-line adjustable "aggressive" stepsize policy [8]; up to this policy, this is nothing but SMP with P_z specified as the unit mass sitting at z, $z \in Z$;
- next, it was solved by SMP (10 runs) with $k_x = 1$, $k_y = 100$ [6] and the stepsize policy

$$\gamma_\tau = \alpha\min\left[\frac{1}{\sqrt{3}\mathcal{L}}, \frac{\sqrt{\Omega_X^2 + \Omega_Y^2}}{\sqrt{7}\sigma\sqrt{\tau}}\right],\ \tau = 1, 2, ...$$

with $\mathcal{L}$ and σ given by (27) (where we replace $\mathbf{V}$ by its valid upper bound 1) and (29) (where we use Ω_X, Ω_Y as given by (36)). When $\alpha = 1$, our stepsize policy becomes the "rolling horizon" version of (11); it can be shown that this policy (which does not require the number t of steps to be chosen in advance) is, theoretically, basically as good as it constant stepsizes prototype). The role of the "acceleration factor" $\alpha \ge 1$ is to allow for larger stepsizes than those given by the worst-case-oriented considerations underlying (11), the option which for DMP is given by the aforementioned on-line adjustable stepsize policy (in our experiments, the latter resulted

[6] with our m, ν, J, the coefficient at k_x in the right hand side of (41) is nearly 30 times larger than the one at k_y, this is why we use $k_y \gg k_x$.

	Iteration count			CPU, sec		
Algorithm	min	mean	max	min	mean	max
DMP	61			2167		
SMP	251	281	351	496	571	708

Table 1: . Effect of randomization, problem (33) ($I = J = 5000, m = 300, \mu = \nu = 2$). In the table: DMP/SMP – Deterministic/Randomized Mirror Prox. Data for SMP are obtained in 10 runs of the algorithm. Running times include those needed to check the termination criterion.

in stepsizes which, at average, were ≈ 250 times the "theoretically safe" ones). The value of α we used (1000) was selected empirically in a small series of pilot experiments and was never revised in the main series of experiments α .

• In every experiment, a solution with the duality gap $\leq \epsilon = 0.01$ was sought. Since the duality gap is not directly observable, this goal was achieved as follows. From time to time (specifically, after every 30 iterations for DMP and every 50 iterations for SMP) we computed $F(z^t)$ for the current approximate solution $z^t = (x^t, y^t)$ (see (9)), thus getting $g := \nabla_x \phi(x^t, y^t)$ and $\mathcal{A}(x^t) = \nabla_y \phi(x^t, y^t)$. We the compute the maximal eigenvalue $\phi^+ = \lambda_{\max}(\mathcal{A}(x^t)$, which is nothing but $\overline{\phi}(x^t) = \max_{y \in Y} \phi(x, y)$, and the quantity $\phi^- = \min_{x \in X}[\phi(x^t, y^t) + \mathrm{Tr}([x - x^t]^T g)]$, which is a lower bound on $\underline{\phi}(y^t) = \min_{x \in X} \phi(x, y^t)$. The quantity $\Delta = \phi^+ - \phi^-$ is an upper bound on $\mathrm{DualityGap}(x^t, y^t)$, and the relation $\Delta \leq \epsilon = 0.01$ was used as the termination criterion.

The results of a typical experiment are presented in table 1. We see that while randomization increases essentially the iteration count, it results in overall reduction of the CPU time by a quite significant factor. It makes sense to note that of 2167 sec CPU time for DMP, 91% (1982 sec) were spent on matrix-vectors multiplications, and just 9% – on computing prox-mappings; for SMP, both these components take nearly equal times.

5.2 Illustration II: low dimensional approximation

Consider the problem as follows: we are given n *unit* vectors $a_j \in \mathbf{R}^m$, $1 \leq j \leq n$, and know that for some given k, $1 < k \leq m/2$, and $\delta \in (0, 1)$ all a_i's are at the $\|\cdot\|_2$-distance at most $\delta < 1$ form certain k-dimensional subspace L, common for all points. The problem is to recover this subspace[7], which reduces to solving the problem

$$\mathrm{Opt}_* = \max_{x \in \mathcal{P}_k} \min_{y \in Y} \sum_{j=1}^{n} y_j a_i^T x a_i, \tag{43}$$

where $\mathcal{P}_k \subset E_x = \mathbf{S}^m$ is the family of all orthoprojectors of rank k on $\mathbf{R}^p$, and $Y = \{y \in \mathbf{R}^n_+ : \sum_j y_j = 1\}$ is the standard simplex in $E_y = \mathbf{R}^m$. The set $\mathcal{P}_k$ is nonconvex; we relax it to the set

$$X = \{x \in \mathbf{S}^m : I_m \succeq x \succeq 0, \mathrm{Tr}(x) = k\},$$

thus arriving at the relaxed saddle point problem

$$\begin{array}{c} -\mathrm{Opt} = \min_{x \in X} \max_{y \in Y} [\phi(x, y) := -\sum_{j=1}^n y_j a_j^T x a_j] \\ F_x(x, y) = -\sum_{j=1}^n y_j a_j a_j^T, \quad F_y(x, y) = [a_1^T x a_1; ...; a_n^T x a_n] \end{array} \tag{44}$$

[7]Note the difference with the PCA – Principal Component Analysis: we want to minimize the maximal, over i, deviation of a_i, from L rather than the sum of squares of these deviations.

(we have equivalently transformed the relaxed problem to fit our standard notation). Note that ϕ is a polynomial of degree $d = 2$ (just bilinear). Let us apply to 44 our approach.

Scale factor. We clearly have $\mathbf{V} \leq 1$ (recall that $\|a_i\|_2 = 1$, $0 \prec x \prec I_m$ for $x \in X$, and $\|y\|_1 \leq 1$ for $y \in Y$).

Setup. We set

$$\begin{array}{rcl} \omega_X(x) & = & \frac{8}{q(1+q)} \sum_{i=1}^m \lambda_i^{1+q}(x),\ q = \min[1, \ln(k)/\ln(m/k)], \\ \omega_Y(y) & = & \frac{8\sqrt{\mathrm{e}}}{p(1+p)} \sum_{j=1}^n y_j^{1+p},\ p = 1/(2\ln(n)), \end{array} \tag{45}$$

thus getting d.-g.f.'s for X, Y compatible with $\|\cdot\|_X$, $\|\cdot\|_Y$, respectively (Proposition A.2 and Remark A.1), the corresponding radii of X, Y are

$$\Omega_x \leq O(1)\sqrt{k\ln(k)/\ln(m/k)},\ \Omega_Y \leq O(1)\sqrt{\ln(n)}, \tag{46}$$

see (67).

Deterministic algorithm. When solving 44 within accuracy $\epsilon < 1$ by the deterministic algorithm DMP,
— the iteration count is $N_{\mathrm{d}}(\epsilon) = O(1)\frac{k\ln(k)/\ln(m/k)+\ln(n)}{\epsilon}$,
— the complexity of an iteration is $O(1)(m^3 + n)$ a.o. for computing prox-mappings and $O(1)m^2 n$ a.o. for computing the values of F.
Note that *as far as deterministic solution algorithms are concerned*, the outlined bounds result in the best known to us overall arithmetic complexity of finding an ϵ-solution in the large scale case.

When $n \gg m$, the cost of prox-mapping is much smaller than the one of computing the values of F, implying that there might be room for accelerating by randomization.

Randomization. In order to compute, given $z = (x, y) \in X \times Y$, unbiased random estimates of $F_x(x, y)$ and $F_y(y)$, we act as follows.

1. We associate with y the distribution P_y on Y as follows: $\eta \sim P_y$ takes the values e_j (basic orths in $\mathbf{R}^n$ with probabilities y_j, $1 \leq j \leq n$ (cf. Example 1); the corresponding random estimate G^x of $F_x(x, y)$ takes the values $-a_j a_j^T$ with probabilities y_j, $1 \leq j \leq n$. Generating the estimate requires the "setup cost" of $O(n)$ a.o.; after this cost is paid, generating of the estimate takes $O(1)[\ln(n) + m^2]$ a.o.

2. We associate with $x \in X$ the distribution P_x on X as follows. Given x, we compute its eigenvalue decomposition $x = U\mathrm{Diag}\{\xi\}U^T$. The vector ξ belongs to the polytope $Q = \{\xi \in \mathbf{R}^m :\leq \xi_i \leq 1, \sum_i \xi_i = k\}$. Now, there is a simple algorithm [7, section A.1] which allows, given $\xi \in Q$, to represent ξ as a convex combination $\sum_{i=1}^m \lambda_i \xi^i$ of extreme points of Q (which are Boolean vectors with exactly k entries equal to 1); the cost of building this representation is $O(1)km^2$ a.o. We build this representation; and define P_x as the distribution of a random symmetric matrix which takes values $U\mathrm{Diag}\{\xi^i\}U^T$ with probabilities λ_i, $1 \leq i \leq m$, so that the random estimate of $F_y(x, y)$ is the vector with the entries $G_j^y = \sum_{\ell \in I_i} (a_i^T \mathrm{Col}_\ell[U])^2$, $1 \leq j \leq n$, where I_i is the set of indexes of the k nonzero

entries of the Boolean vector ξ^i, and i takes values $1, ..., m$ with probabilities $\lambda_1, ..., \lambda_m$. Finally, we set $P_z = P_x \times P_y$. Note that this distribution is supported on $X \times Y$ (i.e., Assumption B is satisfied with $\rho = 0$). The "setup" cost of sampling from P_x is $O(1)m^3$ a.o.; after this cost is paid, generating a sample value of G^y costs $O(1)kmn$ a.o.

With the outlined randomization, the cost of generating a sample value of G_{k_x,k_y} in the range $\ln(n) \le O(1)m^2$ costs

$$O(1)(m^3 + k_x kmn + k_y m^2) \text{ a.o.}$$

When $n \gg m \gg k$ and k_x, k_y are moderate, this cost is by far less than the cost $O(1)m^2 n$ of deterministic computation of $F(x, y)$, so that our randomization indeed possesses some potential. Analysis completely similar to the one in section 5.1 shows that our current situation is completely similar to the one in the latter section: while with $k_x = O(1)$, $k_y = O(1)$, the iteration count for the randomized algorithm is proportional to ϵ^{-2} instead of being proportional to ϵ^{-1}, as for the deterministic algorithm, the growth in this count, in certain meaningful range of values of k, m, n, ϵ is by far overweight by reduction in the cost of an iteration. As a result, for ϵ fixed and in the case of appropriate proportion between k, m, n, the randomized algorithm progressively outperforms its deterministic competitor as the sizes of the problem grow.

Numerical illustration. In the experiment we are about to describe, the sizes of problem 44 were selected as

$$m = 100,\ k = 10,\ n = 300,000.$$

The data points a_j were selected at random in certain "smart" way aimed at creating difficult instances; we are not sure that this goal was indeed achieved, but at least the PCA solution (which, with the straightforward random generation of a_j, turns out to recover perfectly well the approximating subspace) was "cut off:" – the largest, over all j, distance of a_j's to the $k = 10$-dimensional PCA subspace in our experiments was as large as 0.99.

Implementation of the approach was completely similar to the one outlined in section 5.1; the only specific issue which should be addressed here is the one of termination. Problem 44 by its origin is no more than a relaxation of the "true" problem (43), so solving it within a given accuracy is of no much interest. Instead, we from time to time (namely, every 10 iterations) took the x-component x^t of the current approximate solution, subject it to eigenvalue decomposition and checked straightforwardly what is the largest, over $j \le n$, $\|\cdot\|_2$-deviation D of a_j from the k-dimensional subspace of $\mathbf{R}^m$ spanned by k principal eigenvectors of x^t. We terminated the solution process when this distance was $\le \delta + \epsilon$, where ϵ is a prescribed tolerance.

Typical experimental results are presented in table 2. The results look surprisingly good – the iteration count is quite low and are the same for both deterministic and randomized algorithms. We do not know whether this unexpected phenomenon reflects the intrinsic simplicity of the problem, or our inability to generate really difficult instances, or the fact that we worked with although reasonable, but not "really small" values of ϵ; this being said, we again see that randomization reduces the CPU time by a quite significant factor.

References

[1] Arora, S., Kale, S., A combinatorial, primal-dual approach to semidefinite programs – in: D. Johnson and U. Feige, Eds., *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 2007* ACM, 2007, 227-236.

	Method	# of steps	CPU, sec	Final deviation D
$\delta = 0.4, \delta + \epsilon = 0.45$	DMP	20	478	0.401
	SMP	20	104	0.427
$\delta = 0.6, \delta + \epsilon = 0.65$	DMP	20	504	0.603
	SMP	20	105	0.620
$\delta = 0.8, \delta + \epsilon = 0.85$	DMP	20	478	0.809
	SMP	20	92	0.819

Table 2: Deterministic (DMP) and randomized (SMP, $k_x = 1$, $k_y = 10$) algorithms on the low dimensional approximation problem.

[2] Baes, M., Buergisser, M., Nemirovski, A., Randomized Mirror-Prox method for solving structured large-scale matrix saddle-point problems – accepted, pending minor revision, to *SIAM Journal on Optimization*
E-print: http://www.arxiv.org/PS_cache/arxiv/pdf/1112/1112.1274v1.pdf

[3] Grigoriadis, M.D., Khachiyan, L.G., A Sublinear-Time Randomized Approximation Algorithm for Matrix Games – *Operations Research Letters* **18** (1995), 53–58.

[4] Juditsky, A., Nemirovski, A., Tauvel, C. Solving variational inequalities with Stochastic Mirror-Prox algorithm – Stochastic Systems 1:1 (2011), DOI: 10.1214/10-SSY011, 17–58.

[5] Judistky, A., Nemirovski, A. (2008) *Large Deviations of Vector-Valued Martingales in 2-Smooth Normed Spaces*. E-print: – `http://arxiv.org/pdf/0809.0813.pdf`

[6] Juditsky, A., Nemirovski, A., First Order Methods for Nonsmooth Large-Scale Convex Minimization, I: General Purpose Methods, II: Utilizing Problem's Structure – in: S. Sra, S. Nowozin, S. Wright, Eds., *Optimization for Machine Learning*, The MIT Press, 2012, 121-184.

[7] Juditsky, A., Kilinç Karzan, F., Nemirovski, A., Randomized First Order Algorithms with Applications to ℓ_1-minimization – *Mathematical Programming*

[8] Nemirovski, A., Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems – *SIAM Journal on Optimization* **15** (2004), 229–251.

[9] Nemirovski, A., Juditsky, A. Lan, G., Shapiro, A., Stochastic Approximation Approach to Stochastic Programming – *SIAM Journal on Optimization* **19:4** (2009), 1574-1609.

[10] Nesterov, Yu., "A Method for Solving a Convex Programming Problem with Rate of Convergence $O(1/k^2)$" – *Soviet Math. Doklady* **27:2** (1983), 372-376.

[11] Nesterov, Yu., "Smooth Minimization of Non-Smooth Functions" – CORE Discussion Paper 2003/12 (2003), *Mathematical Programming* **103** (2005), 127–152.

A Proofs

A.1 Proof of Lemma 4.1

In what follows, C_i are positive quantities depending solely on d, and $\widehat{Z}$ is the convex hull of $\{0\} \cup Z$. Observe that $L[\widehat{Z}] = \mathrm{Lin}(\widehat{Z} - \widehat{Z}) \supset L[Z]$ and $\widehat{Z}^s := \frac{1}{2}[\widehat{Z} - \widehat{Z}] \supset Z^s$; as a result,

$$\|z\|_{\widehat{Z}} \leq \|z\| \ \forall z \in L[Z]. \tag{47}$$

1^0. Observe that for some C_1 one has

$$\forall (z \in \widehat{Z}, 2 \leq k \leq d) : |Q_k(z, ..., z)| \leq C_1 \mathbf{V}. \tag{48}$$

Indeed, let $z \in \widehat{Z}$. The univariate polynomial

$$p(t) := \widehat{\phi}(tz) = \sum_{k=2}^{d} Q_k(z, ..., z) t^k$$

on the segment $0 \leq t \leq 1$ is bounded in absolute value by $\mathbf{V}$ (since $\mathbf{V}$ is the variation of $\widehat{\phi}$ on $\widehat{Z} \ni 0$ and $\widehat{\phi}(0) = 0$), so that the moduli $|Q_k(z, ..., z)|$ of its coefficients are bounded by $C_1 \mathbf{V}$ for some C_1 depending solely on d.

2^0. Our next observation is that for some C_2 one has

$$\forall (z \in L[\widehat{Z}], 2 \leq k \leq d) : |Q_k(z, ..., z)| \leq C_2 \mathbf{V} \|z\|_{\widehat{Z}}^k. \tag{49}$$

Indeed, let $2 \leq k \leq d$. By homogeneity it suffices to verify (49) when $\|z\|_{\widehat{Z}} = 1$, so that $z = \frac{1}{2}[z^1 - z^2]$ with some $z^1, z^2 \in \widehat{Z}$. Setting $h(t_1, t_2) = t_1 z^1 + t_2 z^2$, consider the polynomial of two variables

$$p(t_1, t_2) = Q_k(h(t_1, t_2), h(t_1, t_2), ..., h(t_1, t_2)).$$

p is a polynomial of degree $\leq k \leq d$ on the 2D plane which is bounded in absolute value by $C_1 \mathbf{V}$ in the triangle $t_1, t_2 \geq 0, t_1 + t_2 \leq 1$ (by (48) combined with the fact that for the outlined t_1, t_2 we have $h(t_1, t_2) = (1 - t_1 - t_2) \cdot 0 + t_1 z^1 + t_2 z^2 \in \widehat{Z}$). As a result, the moduli of the coefficients of p do not exceed $C_3 \mathbf{V}$ with appropriately chosen C_3, whence $p(1/2, -1/2) = Q_k(z, ..., z)$ is bounded in absolute value by $C_2 \mathbf{V}$ with appropriately chosen C_2.

3^0. Now let $2 \leq k \leq d$, and let $z^1, ..., z^k \in L[\widehat{Z}]$, $\|z^i\|_{\widehat{Z}} \leq 1$, $1 \leq i \leq k$. Consider the polynomial of k real variables

$$p(t_1, ..., t_k) = Q_k(\sum_{i=1}^{k} t_i z^i, \sum_{i=1}^{k} t_i z^i, ..., \sum_{i=1}^{k} t_i z^i).$$

The degree of this polynomial does not exceed $k \leq d$, and

$$|p(t_1, ..., t_k)| \leq C_2 \mathbf{V} \|t_1 z^1 + ... + t_k z^k\|_{\widehat{Z}}^k \leq C_2 \mathbf{V} \|t\|_1^k$$

by (49). It follows that for some C_4 we have

$$\left| \frac{\partial^k p(t_1, ..., t_k)}{\partial t_k \partial t_{k-1} ... \partial t_1} \right| \leq C_4 \mathbf{V}$$

The left hand side in this relation is $k!|Q_k(z^1,...,z^k)|$ (recall that $Q_k(\cdot,...,\cdot)$ is k-linear and symmetric), and we see that

$$\forall\{z^i\in L[\widehat{Z}],\|z^i\|_{\widehat{Z}}\le 1\}_{i=1}^k : |Q_k(z^1,...,z^k)|\le\frac{C_4}{k!}\mathbf{V},$$

which by homogeneity implies (26).

4^0. It remains to prove the "in particular" part of Lemma 4.1. Taking into account (23), (20) — (22), to this end it suffices to verify that the second order directional derivative $D^2\phi(z)[h,h]=\frac{d^2}{dt^2}\big|_{t=0}\phi(z+th)$ taken at a point $z\in Z$ along a direction $h\in L[Z]$ satisfies

$$|D^2\phi(z)[h,h]|\le\mathcal{L}\|h\|^2$$

with $\mathcal{L}$ given by (27). This is immediate: by (2) we have

$$D^2\phi(z)[h,h]=\sum_{k=2}^d k(k-1)Q_k(h,h,z,...,z).$$

We have $\|z\|_{\widehat{Z}}\le 2$ by definition of $\|\cdot\|_{\widehat{Z}}$ (recall that $z\in Z$), so that by (26) the modulus of the right hand side does not exceed $\sum_{k=2}^d k(k-1)2^{k-2}\|h\|_{\widehat{Z}}^2C^{(1)}\mathbf{V}$. It remains to note that $\|h\|_{\widehat{Z}}\le\|h\|$ due to $h\in L[Z]$ and (47). □

A.2 Proof of Lemma 4.2

1^0. Let, as always, $\widehat{Z}$ be the convex hull of $\{0\}\cup Z$, and let us fix $z\in Z$. Consider the random vectors ζ_x,ζ_y taking values in E_x, E_y, respectively:

$$\begin{array}{l}\zeta_x=G_x[z^1,...,z^{d-1}],\ \zeta_y=G_y[z^1,...,z^{d-1}],\ \zeta=[\zeta_x;\zeta_y],\\ \delta_x=\zeta_x-F_x(z),\ \delta_y=\zeta_y-F_y(z),\ \delta=[\delta_x;\delta_y],\end{array}$$

where $z^1,...,z^{d-1}$ are drawn, independently of each other, from P_z. We claim that for some C_5, depending solely on d, it holds

$$\|\delta\|_*\le C_5\mathbf{V}(1+\rho)^{d-1}. \tag{50}$$

Indeed, by construction of $G[z^1,...,z^{d-1}]$ and in view of (2) we have

$$\begin{array}{rl}&\forall h\in L[Z]:\left\{\begin{array}{rcl}\langle\zeta,h\rangle&=&\sum_{k=1}^d kQ_k(Dh,z^1,...,z^{k-1})\\ \langle F(z),h\rangle&=&\sum_{k=1}^d kQ_k(Dh,z,...,z)\end{array}\right.\\ \Rightarrow\ &\|\delta\|_*=\max\limits_{h\in L[Z],\|h\|\le1}\sum_{k=1}^d k[Q_k(Dh,z,...,z)-Q_k(Dh,z^1,...,z^{k-1})]\\ &\le\max\limits_{h:\|h\|\le1}\sum_{k=2}^d kC^{(1)}\mathbf{V}\|Dh\|_{\widehat{Z}}[\|z\|_{\widehat{Z}}^{k-1}+\|z^1\|_{\widehat{Z}}\|z^2\|_{\widehat{Z}}...\|z^{k-1}\|_{\widehat{Z}}]\end{array}\tag{51}$$

where the concluding inequality is due to (26) (take into account that $h\in L[Z]=L[X]\times L[Y]$, whence $Dh\in L[Z]\subset L[\widehat{Z}]$, and that $z,z^i\in\mathrm{Aff}(Z)\subset L[\widehat{Z}]$). Invoking (47), we get $\|Dh\|_{\widehat{Z}}\le\|Dh\|=\|h\|$. Besides this, $z\in Z$ implies that $\|z\|_{\widehat{Z}}\le 2$, while Assumption B combines with (47) and the relation $\|z'\|_{\widehat{Z}}\le 2$ for all $z'\in Z$ to imply that $\|z^i\|_{\widehat{Z}}\le 2(1+\rho)$. In view of these observations, the concluding quantity in (51) is $\le\sum_{k=2}^d kC^{(1)}\mathbf{V}2^{k-1}[1+(1+\rho)^{k-1}]$, so that $\|\delta\|_*\le C_5\mathbf{V}(1+\rho)^{d-1}$ with C_5 depending solely on d, as claimed in (50). □

2^0. We need the following fact:

Proposition A.1 *Let F be a Euclidean space, $\|\cdot\|$ be a norm on F, $\|\cdot\|_*$ be the conjugate norm, let Ξ be a Polish space equipped with a Borel probability distribution, and $\mathcal{F}$ be the space of all Borel mappings $f:\Xi\to F$ such that for some $c_f\in(0,\infty)$ it holds $\mathbf{E}\{\exp\{\|f(\cdot)\|_*^2/c_f^2\}\}\le\exp\{1\}$. Then*

(i) *$\mathcal{F}$ is a linear space, and the quantity $\sigma[f]=\inf\{c>0:\mathbf{E}\{\exp\{\|f(\cdot)\|_*^2/c^2\}\}\le\exp\{1\}\}$ is a (semi)norm on $\mathcal{F}$;*

(ii) *Let U be a convex compact set in F such that $U^s=\frac{1}{2}[U-U]$ is the unit ball of the norm $\|\cdot\|$. Assume that U admits a d.-g.f. $\omega(\cdot)$ compatible with $\|\cdot\|$, and let Ω be the ω-radius of U^s. Then for properly chosen absolute constant $O(1)$, with $\chi=O(1)\Omega$ the following holds true:*

> *(!) Let $f_1,f_2,...$ be an F-valued martingale-difference, that is, a sequence of random vectors taking values in F and such that $\mathbf{E}_{|t-1}\{f_t\}\equiv 0$ for all t, where $\mathbf{E}_{|t-1}$ is the conditional expectation w.r.t. taken w.r.t. the σ-algebra spanned by $f_1,...,f_{t-1}$. Assume that for a sequence of nonnegative deterministic reals $\sigma_1,\sigma_2,...$ it holds*
>
> $$\mathbf{E}_{|t-1}\{\exp\{f_t(\cdot)\|_*^2/\sigma_t^2\}\}\le\exp\{1\}\ \ a.s.$$
>
> *Then for every t one has*
>
> $$\sigma[f_1+...+f_t]\le\chi\sqrt{\sum\nolimits_{\tau=1}^t\sigma_\tau^2}. \tag{52}$$

Proof. (i) is well known; for the sake of completeness, here is the proof. The fact which indeed needs verification is the triangle inequality. Thus, let $f,g\in\mathcal{F}$, $a>\sigma[f]$ and $b>\sigma[g]$; all we need is to prove that $a+b\ge\sigma[f+g]$. Setting $\lambda=a/(a+b)$, we have

$$\begin{array}{l}\exp\{\|f+g\|_*^2/(a+b)^2\}\le\exp\{[\|f\|_*+\|g\|_*]^2/(a+b)^2\}\\ =\exp\{[\lambda(\|f\|_*/a)+(1-\lambda)(\|g\|_*/b)]^2\}\le\lambda\exp\{(\|f\|_*/a)^2\}+(1-\lambda)\exp\{(\|g\|_*/b)^2\},\end{array}$$

where the concluding $\le$ is due to the convexity of the univariate function $\exp\{s^2\}$. Taking expectations in the resulting inequality, we get $\mathbf{E}\{\exp\{\|f+g\|_*^2/(a+b)^2\}\le\exp\{1\}$, that is, $a+b\ge\sigma[f+g]$, as claimed. (i) is justified.

(ii): Let $\psi(u)=\omega(u/2):\frac{1}{2}U\to\mathbf{R}$, and let $f(\xi)=\max_{u\in\frac{1}{2}U}[\langle\xi,u\rangle-\psi(u)]$ be the Fenchel transform of ψ. Since ω is strongly convex, modulus 1, w.r.t. $\|\cdot\|$, ψ is strongly convex, modulus $1/4$, w.r.t. $\|\cdot\|$, whence, by the standard properties of the Fenchel transformation, f possesses Lipschitz continuous gradient, specifically, $\|f'(\xi)-f'(\eta)\|\le4\|\xi-\eta\|_*$ for all ξ,η. The Fenchel transform of the function $\psi_-(u)=\psi(-u)$ is $f_-(\xi)=f(-\xi)$. Now let ψ^s be the inf-convolution of $\psi(\cdot)$ and $\psi_-(\cdot)$, i.e., the function

$$\begin{array}{rcl}\psi^s(u)&=&\inf_{v,w:v+w=u}(\psi(v)+\psi_-(w))=\inf_{v,w':v-w'=u}(\psi(v)+\psi(w'))\\ &=&\left\{\begin{array}{ll}\theta(u):=\min_{v,w'\in\frac{1}{2}U:v-w'=u}[\psi(v)+\psi(v')], & u\in U^s=\frac{1}{2}[U-U]\\ +\infty, & u\notin U^s\end{array}\right.\end{array}$$

The Fenchel transform of the inf-convolution of ψ and ψ_- is the sum of the Fenchel transforms of ψ and ψ_- (recall that the functions are convex with closed compact domains and are continuous on their domains), that is, it is the function $g(\xi)=f(\xi)+f(-\xi)$. In particular, the Fenchel transform of $\theta(\cdot)$ satisfies $\|g'(\xi)-g'(\eta)\|\le8\|\xi-\eta\|_*$. By the standard properties of the Fenchel

transform, it follows that $\theta(\cdot)$ is strongly continuous, modulus $\frac{1}{8}$, on its domain (which is exactly the unit ball U^s of the norm $\|\cdot\|$, and the variation (the maximum minus the minimum) of θ on the domain is Ω^2 (since the variation $\psi(\cdot)$ over $\frac{1}{2}U$, that is, the variation of $\omega(\cdot)$ over U, is $\Omega^2/2$. The bottom line is that the unit ball U^s of $\|\cdot\|$ admits a continuous strongly convex, modulus 1 w.r.t. $\|\cdot\|$, function (specifically, $8\theta(\cdot)$) with variation over U^s not exceeding $8\Omega^2$. Invoking [5, Proposition 3.3], it follows that the space $(F, \|\cdot\|_*)$ is $O(1)\Omega^2$ regular (for details, see [5]). With this in mind, the conclusion (!) in (ii) is an immediate consequence of [5, Theorem 2.1.(ii)]. □

3^0. Now we can complete the proof of Lemma 4.2. We have already seen that $\mathcal{SO}$ generates unbiased random estimates of F, whence $\mathcal{SO}_{k_x,k_y}$ possesses the same property; thus, $\mathcal{SO}_{k_x,k_y}$ meets the requirement $(10.b)$, which is the first claim in Lemma 4.2. Now let us prove the second claim in this Lemma. In the notation from item 1^0, setting $F = L[X]$ and denoting by π the orthoprojector of E_x onto $F \subset E_x$, (50) implies that

$$\|\pi\delta_x\|_{X,*} = \|\delta_x\|_{X,*} \le C_5\mathbf{V}(1+\rho)^{d-1} \tag{53}$$

(since $\|\delta\|_* = \|\delta_x\|_{X,*} + \|\delta_y\|_{Y,*}$). The x-component Δ_x of the "observation error" of $\mathcal{SO}_{k_x,k_y}$ (the difference $\Delta = [\Delta_x; \Delta_y]$ of the random estimate of $F(z)$ generated by $\mathcal{SO}_{k_x,k_y}$ and $F(z)$) is

$$\Delta_x = \sum_{t=1}^{k_x} f_t \Rightarrow \pi\Delta_x = \sum_{t=1}^{k_x} \widetilde{f}_t, \tag{54}$$

where $\widetilde{f}_1, ..., \widetilde{f}_{k_x}$ are independent copies of the zero mean random vector $k_x^{-1}\pi\delta_x \in F$. Besides this, choosing a point $\bar{x} \in X$ and setting $\widetilde{X} = X - \bar{x} \subset F$, $\widetilde{\omega}(\xi) = \omega(\bar{x}+\xi)$, $\xi \in \widetilde{X}$, we see that $X^s = \frac{1}{2}[\widetilde{X} - \widetilde{X}]$ admits a d.-g-f., specifically, $\widetilde{\omega}(\cdot)$, which is compatible with $\|\cdot\|_X$ and is such that the $\widetilde{\omega}$-radius of $\widetilde{X}$ is Ω_X. Invoking Proposition A.1.(ii) and taking into account that we are in the situation $\sigma[\widetilde{f}_t] = \sigma[f_t] \le C_5 k_x^{-1}\mathbf{V}(1+\rho)^{d-1}$ by (53), we get that for properly chosen C_6 depending solely on d we have

$$\mathbf{E}\left\{\exp\{\|\Delta_x\|_{X,*}^2/\widetilde{\sigma}_x^2\}\right\} \le \exp\{1\},\ \widetilde{\sigma}_x = C_6\Omega_X\mathbf{V}(1+\rho)^{d-1}/\sqrt{k_x}$$

(note that $\|\Delta_x\|_{X,*} = \|\pi\Delta_x\|_{X,*}\|$). Besides this, by (53) $\|\widetilde{f}_t\|_{X,*} \le C_5 k_x^{-1}\mathbf{V}(1+\rho)^{d-1}$ almost surely, whence

$$\mathbf{E}\left\{\exp\{\|\Delta_x\|_{X,*}^2/\bar{\sigma}_x^2\}\right\} \le \exp\{1\},\ \bar{\sigma}_x = C_5\mathbf{V}(1+\rho)^{d-1}.$$

The bottom line is that with properly selected C_7 depending solely on d and with

$$\sigma_x = C_7\mathbf{V}(1+\rho)^{d-1}\min[1, \Omega_X/\sqrt{k_x}]$$

we have

$$\mathbf{E}\left\{\exp\{\|\Delta_x\|_{X,*}^2/\sigma_x^2\}\right\} \le \exp\{1\}.$$

By similar reasons, with properly selected C_8 depending solely on d and with

$$\sigma_y = C_8\mathbf{V}(1+\rho)^{d-1}\min[1, \Omega_Y/\sqrt{k_y}]$$

we have

$$\mathbf{E}\left\{\exp\{\|\Delta_y\|_{Y,*}^2/\sigma_y^2\}\right\} \le \exp\{1\}.$$

Taking into account that $\|\Delta\|_* = \|\Delta_x\|_{X,*} + \|\Delta_y\|_{Y,*}$ and item (i) of Proposition A.1, the second claim in Lemma 4.2 follows. □

A.3 Proofs for section 5

A. Let $\mathbf{S}^m$ be the space of $m \times m$ symmetric matrices equipped with the Frobenius inner product; for $y \in \mathbf{S}^m$, let $\lambda(y)$ be the vector of eigenvalues of y (taken with their multiplicities in the non-ascending order). For an integer k, $1 \le k \le m$, let $Y^k = \{y \in \mathbf{S}^m : \|\lambda(y)\|_\infty \le 1, \|\lambda(y)\|_1 \le k\}$, so that Y^k is the unit ball of certain rotation-invariant norm $\|\cdot\|_{(k)}$ on $\mathbf{S}^m$.

Lemma A.1 *Let m, n, k be integers such that $m \ge n \ge k \ge 1$, and let F be a linear subspace in $\mathbf{S}^m$ such that every matrix $y \in F$ has at most n nonzero eigenvalues. Let, further, $q \in (0,1)$, and let*

$$\chi(y) = \frac{1}{1+q}\sum_{j=1}^m |\lambda_j(y)|^{1+q} : \mathbf{S}^m \to \mathbf{R}.$$

The function $\chi(\cdot)$ is continuously differentiable, convex, and its restriction on the set $Y_F^k = \{y \in F : \|y\|_{(k)} \le 1\}$ is strongly convex w.r.t. $\|\cdot\|_{(k)}$ modulus

$$\beta = q\min[1, \frac{1}{2}k^{1+q}n^{-q}]. \tag{55}$$

Proof. 1^0. Observe that

$$\chi(y) = \mathrm{Tr}(f(y)),\ f(s) = \frac{1}{1+q}|s|^{1+q}. \tag{56}$$

Function $f(s)$ is continuously differentiable on the axis and twice continuously differentiable outside of the origin; consequently, we can find a sequence of polynomials $f_r(s)$ converging, as $r \to \infty$, to f along with their first derivatives uniformly on every compact subset of $\mathbf{R}$ and, besides this, converging to f uniformly along with the first and the second derivative on every compact subset of $\mathbf{R}\backslash\{0\}$. Now let $y, h \in \mathbf{S}^m$, let $y = u\mathrm{Diag}\{\lambda\}u^T$ be the eigenvalue decomposition of y, and let $h = u\widehat{h}u^T$. For a polynomial $p(s) = \sum_{\ell=0}^L p_\ell s^\ell$, setting $P(w) = \mathrm{Tr}(\sum_{\ell=0}^L p_\ell w^\ell) : \mathbf{S}^m \to \mathbf{R}$, and denoting by γ a closed contour in $\mathbf{C}$ encircling the spectrum of y, we have

$$\begin{array}{ll}
(a) & P(y) = \mathrm{Tr}(p(y)) = \sum_{j=1}^m p(\lambda_j(y)) \\
(b) & DP(y)[h] = \mathrm{Tr}(\sum_{\ell=0}^L \ell p_\ell \mathrm{Tr}(y^{\ell-1}h)) = \mathrm{Tr}(p'(y)h) = \sum_{j=1}^m p'(\lambda_j(y))\widehat{h}_{jj} \\
(c) & D^2P(y)[h,h] = \frac{d}{dt}\big|_{t=0} DP(y+th)[h] = \frac{d}{dt}\big|_{t=0}\mathrm{Tr}(p'(y+th)h) \\
& = \frac{d}{dt}\big|_{t=0}\frac{1}{2\pi\imath}\oint_\gamma \mathrm{Tr}(h(zI-(y+th))^{-1})p'(z)dz = \frac{1}{2\pi\imath}\oint_\gamma \mathrm{Tr}(h(zI-y)^{-1}h(zI-y)^{-1})p'(z)dz \\
& = \frac{1}{2\pi\imath}\oint_\gamma \sum_{i,j=1}^m \widehat{h}_{ij}^2 \frac{p'(z)}{(z-\lambda_i(y))(z-\lambda_j(y))}dz = \sum_{i,j=1}^n \widehat{h}_{ij}^2\Gamma_{ij}, \\
& \Gamma_{ij} = \begin{cases} \frac{p'(\lambda_i(y))-p'(\lambda_j(y))}{\lambda_i(y)-\lambda_j(y)}, & \lambda_i(y) \ne \lambda_j(y) \\ p''(\lambda_i(y)), & \lambda_i(y) = \lambda_j(y) \end{cases}
\end{array}$$

We conclude from (a,b) that as $k \to \infty$, the real-valued polynomials $F_r(\cdot) = \mathrm{Tr}(f_r(\cdot))$ on $\mathbf{S}^m$ converge, along with their first order derivatives, uniformly on every bounded subset of $\mathbf{S}^m$, and the limit of the sequence, by (a), is exactly $\chi(\cdot)$. Thus, $\chi(\cdot)$ is continuously differentiable, and (b) says that

$$D\chi(y)[h] = \sum_{j=1}^m f'(\lambda_j(y))\widehat{h}_{jj}. \tag{57}$$

Besides this, (a-c) say that if U is a closed convex set in $\mathbf{S}^m$ which does not contain singular matrices, then $F_r(\cdot)$, as $r \to \infty$, converge along with the first and the second derivative uniformly on every compact subset of U, so that $\chi(\cdot)$ is twice continuously differentiable on U, and at every point $y \in U$ we have

$$D^2\chi(y)[h,h] = \sum_{i,j=1}^m \widehat{h}_{ij}^2 \Gamma_{ij},\ \Gamma_{ij} = \begin{cases} \frac{f'(\lambda_i(y)) - f'(\lambda_j(y))}{\lambda_i(y) - \lambda_j(y)}, & \lambda_i(y) \neq \lambda_j(y) \\ f''(\lambda_i(y)), & \lambda_i(y) = \lambda_j(y) \end{cases} \tag{58}$$

and in particular $\chi(\cdot)$ is convex on U.

3^0. We intend to prove that (i) $\chi(\cdot)$ is convex, and (ii) its restriction on the set Y_F^k is strongly convex, with certain modulus $\alpha > 0$, w.r.t. the norm $\|\cdot\|_{(k)}$. Since χ is continuously differentiable, all we need to prove (i) is to verify that

$$\langle \chi'(y') - \chi'(y''), y' - y'' \rangle \geq 0 \tag{$*$}$$

for a dense in $\mathbf{S}^m \times \mathbf{S}^m$ set of pairs (y', y''), e.g., those with nonsingular $y' - y''$. For a pair of the latter type, the polynomial $q(t) = \mathrm{Det}(y' + t(y'' - y'))$ of $t \in \mathbf{R}$ is not identically zero and thus has finitely many roots on $[0,1]$. In other words, we can find finitely many points $t_0 = 0 < t_1 < ... < t_n = 1$ such that all "matrix intervals" $\Delta_i = (y_i, y_{i+1})$, $y_k = y' + t_k(y'' - y')$, $1 \leq i \leq n-1$, are comprised of nonsingular matrices. Therefore χ is convex on every closed segment contained in one of Δ_i's, and since χ is continuously differentiable, ($*$) follows.

4^0. It remains to prove that with β given by (55) one has

$$\langle \chi'(y') - \chi'(y''), y' - y'' \rangle \geq \alpha |y' - y''|_1^2 \ \forall y', y'' \in Y_F^k \tag{59}$$

Let $\epsilon > 0$, and let Y^ϵ be a convex open in $Y^k = \{y : \|y\|_{(k)} \leq 1\}$ neighborhood of Y_F^k such that for all $y \in Y^\epsilon$ at most n eigenvalues of y are of magnitude $> \epsilon$. We intend to prove that for some $\alpha_\epsilon > 0$ one has

$$\langle \chi'(y') - \chi'(y''), y' - y'' \rangle \geq \alpha_\epsilon \|y' - y''\|_{(k)}^2 \ \forall y', y'' \in Y^\epsilon. \tag{60}$$

Same as above, it suffices to verify this relation for a dense in $Y^\epsilon \times Y^\epsilon$ set of pairs $y', y'' \in Y^\epsilon$, e.g., for those pairs $y', y'' \in Y^\epsilon$ for which $y' - y''$ is nonsingular. Defining matrix intervals Δ_i as above and taking into account continuous differentiability of χ, it suffices to verify that if $y \in \Delta_i$ and $h = y' - y''$, then $D^2\chi(y)[h,h] \geq \alpha_\epsilon |h|_1^2$. To this end observe that by (58) all we have to prove is that

$$D^2\chi(y)[h,h] = \sum_{i,j=1}^m \widehat{h}_{ij}^2 \Gamma_{ij} \geq \alpha_\epsilon \|h\|_{(k)}^2. \tag{$\#$}$$

5^0. Setting $\lambda_j = \lambda_j(y)$, observe that $\lambda_i \neq 0$ for all i due to the origin of y. We claim that if $|\lambda_i| \geq |\lambda_j|$, then $\Gamma_{ij} \geq q|\lambda_i|^{q-1}$. Indeed, the latter relation definitely holds true when $\lambda_i = \lambda_j$. Now, if λ_i and λ_j are of the same sign, then $\Gamma_{ij} = \frac{|\lambda_i|^q - |\lambda|_j^q}{|\lambda_i| - |\lambda_j|} \geq q|\lambda_i|^{q-1}$, since the derivative of the concave (recall that $0 < q \leq 1$) function t^q of $t > 0$ is positive and nonincreasing. If λ_i and λ_j are of different signs, then $\Gamma_{ij} = \frac{|\lambda_i|^q + |\lambda_j|^q}{|\lambda_i| + |\lambda_j|} \geq |\lambda_i|^{q-1}$ due to $|\lambda_j|^q \geq |\lambda_j||\lambda_i|^{q-1}$, and therefore $\Gamma_{ij} \geq q|\lambda_i|^{q-1}$. Thus, our claim is justified.

W.l.o.g. we can assume that the positive reals $\mu_i = |\lambda_i|$, $i = 1, ..., m$, form a nondecreasing sequence, so that, by above, $\Gamma_{ij} \geq q\mu_j^{q-1}$ when $i \leq j$. Besides this, at most n of μ_j are $\geq \epsilon$,

since $y', y'' \in Y^\epsilon$ and therefore $y \in Y^\epsilon$ by convexity of Y^ϵ. By the above,

$$D^2\chi(y)[h,h] \geq 2q \sum_{i<j\leq m} \widehat{h}_{ij}^2 \mu_j^{q-1} + q \sum_{j=1}^m \widehat{h}_{jj}^2 \mu_j^{q-1},$$

or, equivalently by symmetry of $\widehat{h}$, if

$$h^j = \begin{bmatrix} & & & \widehat{h}_{1j} \\ & & & \widehat{h}_{2j} \\ & & & \vdots \\ \widehat{h}_{j1} & \widehat{h}_{j2} & \cdots & \widehat{h}_{jj} \\ & & & \end{bmatrix}$$

and H_j is the Frobenius norm $\|h^j\|_{\rm Fro}$ of h^j, then

$$D^2\chi(y)[h,h] \geq q \sum_{j=1}^m H_j^2 \mu_j^{q-1}. \tag{61}$$

6^0. Now note that

$$0 < \mu_j \leq 1 \, \forall j, \, \mu_j \leq \epsilon, \, j \leq m-n, \, \sum_{j=1}^m \mu_j \leq k \tag{62}$$

due to $y \in Y^\epsilon \subset Y^k$ and $\mu_j > 0$ for all j. Now, by the definition of $\|\cdot\|_{(k)}$, setting

$$\eta = \|h\|_{(k)} \, [\equiv \|\widehat{h}\|_{(k)}],$$

observe that either η is the spectral norm $\|\lambda(\widehat{h})\|_\infty$ of $\widehat{h}$, or $k\eta$ is the nuclear norm of $\widehat{h}$. In the first case, the Frobenius norm of $\widehat{h}$ is $\geq \eta$, meaning that $\sum_{j=1}^m H_j^2 = \|\widehat{h}\|_{\rm Fro}^2 = \eta^2$. Since $q \in (0,1)$ and $0 < \mu_j \leq 1$ for all j by (62), we conclude from (61) and from the evident relation $\|\widehat{h}\|_{\rm Fro}^2 = \sum_j \|h^j\|_{\rm Fro}^2 = \sum_j H_j^2$ that in the case in question we have

$$D^2\chi(y)[h,h] \geq q \sum_{j=1}^m H_j^2 \geq q\eta^2 \equiv q\|h\|_{(k)}^2. \tag{63}$$

Now assume that we are in the second case:

$$k\|h\|_{(k)} = k\eta = \|h\|_{\rm nuc} = \|\widehat{h}\|_{\rm nuc}. \tag{64}$$

Observe that h^j are matrices of rank ≤ 2, so that $\|h^j\|_{\rm nuc} \leq \sqrt{2} H_j$, and since $\widehat{H} = \sum_{j=1}^m h^j$, we have $\|\widehat{h}\|_{\rm nuc} \leq \sum_j \|h_j\|_{\rm nuc} \leq \sqrt{2} \sum_j H_j$, which combines with (64) to imply the first inequality in the following chain:

$$\begin{array}{l}
k^2\|h\|_{(k)}^2 = \|\widehat{h}\|_{\rm nuc}^2 \leq 2\left(\sum_{j=1}^m H_j\right)^2 = 2\left(\sum_{j=1}^m [H_j \mu_j^{(q-1)/2}] \mu_j^{(1-q)/2}\right)^2 \\
\leq 2\left(\sum_{j=1}^m \mu_j^{q-1} H_j^2\right)\left(\sum_{j=1}^m \mu_j^{1-q}\right) \ \text{[Cauchy inequality]} \\
\leq 2q^{-1} D^2\chi(y)[h,h]\left(\sum_{j=1}^m \mu_j^{1-q}\right) \ \text{[by (61)]} \\
\leq 2q^{-1} D^2\chi(y)[h,h]\left((m-n)\epsilon^{1-q} + \sum_{j=m-n+1}^m \mu_j^{1-q}\right) \ \text{[by (62)]} \\
\leq 2q^{-1} D^2\chi(y)[h,h]\left((m-n)\epsilon^{1-q} + [n^{-1}\sum_{j=m-n+1}^m \mu_j]^{1-q} n\right) \ \text{[since } 0 < q < 1\text{]} \\
\leq 2q^{-1} D^2\chi(y)[h,h]\left((m-n)\epsilon^{1-q} + k^{1-q} n^q\right). \ \text{[by (62)]}
\end{array}$$

Thus, in the case of (64) we have

$$D^2\chi(y)[h,h] \geq \frac{q}{2}\frac{k^2}{(m-n)\epsilon^{1-q}+k^{1-q}n^q}\|h\|_{(k)}^2.$$

Setting

$$\alpha_\epsilon = q\min[1, \frac{1}{2}\frac{k^2}{(m-n)\epsilon^{1-q}+k^{1-q}n^q}] \tag{65}$$

and recalling (63), we arrive at the desired inequality (#).

7^0. As we have already explained, (#) implies the validity of (60) (and therefore – the validity of (59)) with $\alpha = \alpha_\epsilon$. Since $\alpha_\epsilon \to \beta$ as $\epsilon \to +0$, (59) indeed is satisfied with β given by (55). □

B. Lemma A.1 is the key to the two statements as follows.

Proposition A.2 *Let k, m be integers such that $1 \leq k \leq m/2$, and let $X = \{x \in \mathbf{S}^m : x \succeq 0, \mathrm{Tr}(x) = k\}$. The function*

$$\omega(x) = \tfrac{4}{\beta(1+q)}\sum_{j=1}^m |\lambda_j(x)|^{1+q},$$
$$q = \begin{cases} \min[1, \ln(k)/\ln(m/k)], & k > 1 \\ 1/(2\ln(m)), & k = 1 \end{cases}, \ \beta = \begin{cases} 1, & k \geq \sqrt{m} \\ q/2, & 1 < k < \sqrt{m} \\ q/(2\sqrt{e}), & k = 1 \end{cases} \tag{66}$$

is convex continuously differentiable function on E which is strongly convex, modulus 1 w.r.t. $\|\cdot\|_X$, on X and thus is a d.-g.f. for X compatible with $\|\cdot\|_X$. The ω-radius of X satisfies

$$\Omega_X \leq 2\sqrt{\frac{2k}{\beta(1+q)}}. \tag{67}$$

Proof. The only non-evident statement is that ω is strongly convex, modulus 1 w.r.t. $\|\cdot\|_X$, on X, and this is what we are about to prove. Let $\|\cdot\|_{(k)}$ be the norm on $\mathbf{S}^m$ with the unit ball $Y^k = \{y \in \mathbf{S}^m : \|\lambda(y)\|_\infty \leq 1, \|\lambda(y)\|_1 \leq k\}$, and let

$$\chi(x) = \frac{1}{1+q}\sum_{j=1}^m |\lambda_j(x)|^{1+q}.$$

When $k \geq \sqrt{m}$, Y^k contains the unit ball of the Frobenius norm, and consequently $\|\cdot\|_{(k)} \leq \|\cdot\|_{\mathrm{Fro}}$, and $q = 1$, meaning that the function $\chi(\cdot) = \frac{1}{2}\|\cdot\|_{\mathrm{Fro}}^2$ is strongly convex, modulus 1, w.r.t. $\|\cdot\|_{\mathrm{Fro}}$, and therefore is strongly convex, modulus $\beta := 1$, w.r.t. $\|\cdot\|_{(k)} \leq \|\cdot\|_{\mathrm{Fro}}$. Let now $k < \sqrt{m}$. In this case $q \in (0,1)$, and therefore, by Lemma A.1, χ is strongly convex, modulus $\beta := q\min[1, \frac{1}{2}k^{1+q}m^{-q}]$, on Y^k. Note that $\beta = q/2$ when $k > 1$ and $\beta = q/(2\sqrt{\mathrm{e}})$ when $k = 1$.

Now observe that X clearly is contained in Y^k, implying that $\chi(x)$ is strongly convex, modulus β w.r.t. $\|\cdot\|_{(k)}$, on X. At the same time, we claim that the $\|\cdot\|_X$-unit ball $X^s \subset L[X] = \{x \in \mathbf{S}^m : \mathrm{Tr}(x) = 0\}$ contains the set $\{x \in L[X] : \|x\|_{(k)} \leq 1/2$, meaning that $\|\cdot\|_X \leq 2\|\cdot\|_{(k)}$ on $L[X]$, as a result, $\chi(\cdot)$ is strongly convex, modulus $\beta/4$ w.r.t. $\|\cdot\|_X$, on X, so that $\omega(x) = (4/\beta)\chi(x)$ is strongly convex, modulus 1 w.r.t. $\|\cdot\|_X$, on X, and this is exactly what we want to prove. To support our claim, let $x \in L[X]$ be such that $\|x\|_{(k)} \leq 1/2$, and let

$x = U\mathrm{Diag}\{\xi\}U^T$ be the eigenvalue decomposition of x. Since $x \in L[X]$ and $\|x\|_{(k)} \le 1/2$, we have

$$(a): \sum_{j=1}^{m} \xi_j = 0, \quad (b): \ |\xi_j| \le 1/2 \, \forall j \le m, \quad (c): \ 2\alpha := \sum_{j=1}^{m} |\xi_j| \le k/2.$$

Now let us select $\delta_j \ge 0$, $1 \le j \le m$, in such a way that

$$(d): \ \delta_j \le 1/2 - |\xi_j| \, \forall j, \quad (e): \ \sum_j \delta_j = \frac{1}{2}k - \alpha.$$

Such a selection is possible due to $|\xi_j| \le 1/2$ (by (b)) and $\sum_{j=1}^{m}[1/2 - |\xi_j|] = m/2 - 2\alpha \ge k/2 - \alpha$ (see (c) and take into account that $k \le m/2$). Now let $\eta^+ = 2(\xi^+ + \delta)$, $\eta^- = 2(\xi^- + \delta)$, where ξ^+ is the vector with coordinates $\max[\xi_i, o]$, and ξ^- is the vector with coordinates $\max[-\xi_i, 0]$. We have $\eta^\pm \ge 0$ (since $\delta \ge 0$) and $\|\eta^\pm\|_\infty \le 1$ (by (d)). Finally, $\sum_j \xi_j^+ = \sum_j \xi_j^- = \alpha$ by (a) and by the definition of α, whence $\sum_j \eta_j^+ = \sum_j \eta_j^- = 2\sum_j \delta_j + 2\alpha = k$ by (e). These relations imply that the symmetric matrices $x^\pm = U\mathrm{Diag}\{\eta^\pm\}U^T$ belong to X, and by construction $x = \frac{1}{2}[x^+ - x^-]$, so that $x \in X^s$, as claimed. □

Proposition A.3 *Let K, M, N be integers such that $1 \le K \le M \le N$, and let $\|\cdot\|_{(K)}$ be the norm on $\mathbf{R}^{M\times N}$ with the unit ball $X = \{x \in \mathbf{R}^{M\times N} : \|\sigma(x)\|_\infty \le 1, \|\sigma(x)\|_1 \le K\}$. Then the function*

$$\omega(x) = \frac{4}{q(1+q)} \sum_{i=1}^{M} \sigma_i^{1+q}(x), \ q = \min[1, \ln(2K)/\ln(M/K)], \tag{68}$$

is convex and continuously differentiable, and its restriction on X is strongly convex, modulus 1 w.r.t. $\|\cdot\|_{(K)}$, on X. The ω-radius Ω_X of X satisfies

$$\Omega_X \le 2\sqrt{\frac{2K}{q(1+q)}}. \tag{69}$$

Proof. The only nontrivial claim is that $\omega(\cdot)$ is strongly convex, modulus 1, w.r.t. $\|\cdot\|_{(K)}$. When $q = 1$, i.e., when $\sqrt{2}K \ge M$, X clearly contains the ball $\{x : \|x\|_{\mathrm{Fro}} \le 1/\sqrt{2}\}$, so that $\|\cdot\|_{(K)} \le \sqrt{2}\|\cdot\|_{\mathrm{Fro}}$, and $\omega(x) = \|x\|_{\mathrm{Fro}}^2$ is strongly convex, modulus 2, w.r.t. $\|\cdot\|_{\mathrm{Fro}}$, and thus indeed strongly concave, modulus 1, w.r.t. $\|\cdot\|_{(k)}$. Now let $q < 1$. Let $m = M + N$, $n = 2M$, $k = 2K$, so that $1 < k \le m$, and let $\mathcal{A}(x) = \begin{bmatrix} & x \\ x^t & \end{bmatrix}$ be the linear mapping from $\mathbf{R}^{M\times N}$ into $\mathbf{S}^m$. It is well known that the eigenvalues of $\mathcal{A}(x)$ are the $n = 2M$ reals $\pm\sigma_i(x)$, $1 \le i \le M$, and $m - n$ zeros. Therefore for the norm $\|\cdot\|_{(k)}$ specified in Lemma A.1 it holds

$$\|x\|_{(K)} = \|\mathcal{A}(x)\|_{(k)} \, \forall x \in \mathbf{R}^{M\times N}. \tag{70}$$

By Lemma A.1, the function $\omega^+(y) = \frac{2}{q(1+q)} \sum_{j=1}^{M} |\lambda_j(y)|^{1+q}$ is convex and continuously differentiable on the entire $\mathbf{S}^m$, and its restriction on the set $Y = \{y \in \mathrm{Im}(\mathcal{A}) : \|y\|_{(k)} \le 1\}$ is strongly convex, modulus 1 w.r.t. $\|\cdot\|_{(k)}$, on Y, implying, due to (70), that the function $\omega(x) = \omega^+(\mathcal{A}(x))$ is convex and continuously differentiable on $\mathbf{R}^{M\times N}$, and its restriction on the unit ball X of the norm $\|\cdot\|_{(K)}$ is strongly convex, modulus 1 w.r.t. $\|\cdot\|_{(K)}$, on B^K. □

Remark A.1 *Note that inspecting the proofs, it is easily seen that the results of Propositions A.2, A.3 remain true if when one replaces* $\mathbf{S}^m$ *(resp.,* $\mathbf{R}^{M\times N}$ *with their subspaces comprised of block-diagonal matrices of a given block-diagonal structure. E.g., when* $1 \le k \le m/2$*, the function*

$$\omega(x) = \frac{4}{\beta(1+q)} \sum_{j=1}^{m} x_j^{1+q}$$

with q, β *given by (66) is a d.-g.f. for the set* $X = \{x \in \mathbf{R}^m : 0 \le x_j \le 1 \, \forall j, \sum_{j=1}^m x_j = k\}$ *compatible with the norm* $\|\cdot\|_X$ *with the unit ball* $X^s = \frac{1}{2}[X - X]$ *on the space* $L[X] = \mathrm{Lin}(X - X) = \{x \in \mathbf{R}^m : \sum_j x_j = 0\}$ *(treat* m*-dimensional vectors as diagonals of* $m \times m$ *diagonal matrices).*